



ТЕХНИЧЕСКИ УНИВЕРСИТЕТ – СОФИЯ

Факултет по телекомуникации

Катедра „Технологии и мениджмънт на комуникационни системи“

Маг. Блерим Исмайл Зюлфиу

**МЕТОДИ И ОНЛАЙН ПЛАТФОРМА ЗА ИЗСЛЕДВАНЕ НА
ДИНАМИКАТА И ОЦЕНКА НА ЖИЗНЕСПОСОБНОСТТА
НА ФИРМА ЗА АВТОМАТИЗИРАНО ПРОЕКТИРАНЕ ВЪВ
ВРЕМЕ НА ТЕХНОЛОГИЧЕН ПРЕХОД**

А В Т О Р Е Ф Е Р А Т

**на дисертация за придобиване на образователна и научна степен
"ДОКТОР"**

Област: 5. Технически науки
Професионално направление 5.3 Комуникационна и компютърна техника
Научна специалност "Автоматизация на инженерния труд и системи за
автоматизирано проектиране"

Научен ръководител: доц. д-р инж. Галя Маринова

София, 2026 г.

Дисертацията се обсъжда и насочва за защита от Съвета на катедрата "Технологии и управление на комуникационни системи" към Факултета по телекомуникации на ТУ-София на редовна среща, проведена на 11.06.2026 г.

Публичната защита на дисертацията ще се проведе на 01.10.2026 г. от 15:00 часа в Конферентната зала на БИК на Техническия университет – София на открито заседание на научното жури, определено с нареждане No ОЖ – 5.3 - 63 / 21.07.2026 на ректора на ТУ-София, съставено от:

1. доц. д-р инж. Тихомир Брусев – Председател
2. доц. д-р Анка Цветанова – Научен секретар
3. проф. дн Даниела Борисова
4. проф. дтн Тодор Стоилов
5. доц. д-р инж. Адриана Бороджиева

Рецензенти:

1. проф. дн Даниела Борисова
2. доц. д-р Анка Цветанова

Материалите по защитата са достъпни за заинтересованите в офиса на Факултета по телекомуникации на ТУ-София, блок No1, офис No 1306.

Дисертантът е задочен докторант към катедра „Технологии и мениджмънт на комуникационни системи“ на факултет по телекомуникации. Изследванията по дисертационната разработка са направени от автора, като някои от тях са подкрепени от научноизследователски проекти.

Автор: Маг. Блерим Исмайл Зюлфиу

Заглавие: Методи и онлайн платформа за изследване на динамиката и оценка на жизнеспособността на фирма за автоматизирано проектиране във време на технологичен преход

Тираж: 20 броя

Отпечатано в ИПК на Технически университет – София

I. ОБЩА ХАРАКТЕРИСТИКА НА ДИСЕРТАЦИОННИЯ ТРУД

Актуалност на проблема

Индустрията за автоматизирано проектиране в електрониката (EDA) се превърна в критична част от глобалната верига на доставки на полупроводници. Тя разработва инструменти за автоматизирано проектиране и интелектуална собственост (IP), използвани при проектирането, верификацията и производството на полупроводници. През последните пет десетилетия EDA преживя многобройни сливания и придобивания (M&A), тъй като полупроводниците еволюираха от няколко хиляди транзистора през 80-те години на XX век до няколко милиарда транзистора в проектите от тип система върху чип. Настоящото изследване идентифицира 661 сливания и придобивания (M&A) в сектора EDA през последните петдесет години, което доведе до концентрацията на много малки EDA компании и до формирането на тясна концентрация на Synopsys, Cadence Design Systems и наскоро появилия се монополистичен конкурент Siemens EDA (Mentor Graphics). Така това изследване е важно за всички заинтересовани страни в полупроводниковата индустрия, както и за регулаторните органи, поради нарасналата концентрация на глобалната полупроводникова индустрия в ключовия технологичен сектор.

Цел на дисертацията, основни задачи и методи на изследване

Целта на дисертационния труд е да се разработи съгласуван набор от методи и онлайн изследователска платформа за изучаване на динамиката и оценяване на жизнеспособността на EDA компаниите във времена на технологичен преход. Започвайки от изграждането на надежден и валидиран набор от данни за M&A транзакции в EDA, който осигурява емпиричната основа за всяка следваща стъпка, през разработването на нови методи на машинно обучение за автоматично идентифициране на стратегически M&A модели, след това захранвайки методи за прогнозиране и анализ на жизнеспособността, които предсказват вълните на сливания и придобивания и оценяват дългосрочната устойчивост на EDA компаниите, до прилагането на разработените методи върху водещи компании от сектора на чиповете за ИИ.

За постигане на целта на дисертацията са дефинирани следните задачи:

Задача 1 – Изграждане и валидиране на изчерпателна база с данни за M&A в EDA, обхващаща пълния исторически хоризонт на индустрията, с всички стандартни полета, необходими за последващия анализ: целева компания, придобиващ, година на основаване, дата на придобиване, продължителност на живота и стойност на сделката, когато е налична. Базата с данни трябва да достигне ниво на пълнота, достатъчно за поддържане на количествени методи, и трябва да бъде подложена на документирана процедура за валидиране.

Задача 2 – Проектиране и валидиране на алгоритъма за обратно хибридно клъстериране (Reverse Hybrid Clustering) за идентифициране на M&A модели. Алгоритъмът изпълнява първо DBSCAN за откриване на шум и ядрена структура, а след това K-Means за прецизиране на ядрените области в компактни клъстери; шумът, отстранен в първия етап, се реинтегрира там, където сходството с прецизиран клъстер надхвърля дефиниран праг. Качеството на клъстеризацията се оценява чрез вътрешни метрики за валидация, включително силуетния коефициент, индекса на Дейвис–Боулдин и индекса на Калински–Харабаш.

Задача 3 – Проектиране и валидиране на композитната метрика TechImpactScore за количествено определяне на технологичното въздействие на M&A транзакциите. Метриката комбинира признаци, свързани с патенти, финансови, времеви и стратегически признаци на всяка транзакция; относителното тегло на всеки признак се извежда емпирично чрез анализа на главните компоненти (PCA) върху набора от данни, вместо да бъде фиксирано априори. Валидирането се извършва чрез корелиране на получения резултат с независими данни за стойността на сделките.

Задача 4 – Разработване на рамка за прогнозиране на вълните на сливания и придобивания в сектора EDA, базирана на сингулярен спектрален анализ. Рамката декомпозира времевия ред на историческата M&A активност на трендов, осцилаторен и шумов компонент и произвежда многогодишна прогноза, която може да бъде оценена спрямо задържани периоди.

Задача 5 – Изграждане на рамка за анализ на жизнеспособността и за предсказване жизнеспособността на EDA компаниите. Рамката използва регресията на пропорционалните рискове на Кокс за оценяване на вероятността от придобиване или провал при дадени характеристики на компанията и пазарни условия; прогностичната ефективност се оценява чрез конкордансния индекс.

Задача 6 – Конструирание на модул за мрежов анализ, който картографира екосистемата на придобиванията в EDA с теоретико-графови техники. Модулът изчислява централностите по степен, по посредничество и по близост за придобиващите компании и отчита индекса на Херфиндал–Хиршман като мярка за пазарна концентрация.

Задача 7 – Реализиране на полуавтоматизиран модул за откриване на M&A, базиран на многоагентна архитектура с големи езикови модели и валидиране с човек в цикъла (human-in-the-loop). Модулът използва специализирани агенти за уеб търсене, извличане на съдържание, валидиране и структуриране на данни, с цел да подпомага изследователите при откриването на нови M&A транзакции при ниски оперативни разходи и висока прецизност.

Задача 8 – Проектиране и внедряване на интегрирана онлайн изследователска платформа, която хоства всеки аналитичен компонент, реализиран в Задачи 1 до 7. Платформата следва трислойна архитектура (релационна база данни, бекенд услуга, излагаща REST API, и едностраничен уеб фронтенд) и предоставя RESTful крайни точки за всяка аналитична операция.

Задача 9 – Валидиране на всеки разработен метод чрез стриктно статистическо тестване. Конвейерът за валидация включва еднофакторен дисперсионен анализ (ANOVA) за диференциране на клъстерите, бутстрап кръстосана валидиране за стабилност на резултатите и проучвания с аблация на признаци за анализ на чувствителността на TechImpactScore.

Разработване на полуавтоматизирана многоагентна система за динамично разширяване на базата с данни за M&A. Накрая, предложената аналитична рамка е интегрирана в онлайн изследователската платформа EDA-DYV.

Научна новост

Дисертационният труд представя няколко иновации. Първо, нова хибридна техника за клъстеризация, наречена обратно хибридно клъстериране (Reverse Hybrid Clustering, RHC), свързва добре известния алгоритъм DBSCAN (Density-Based Spatial Clustering of Applications with Noise) и алгоритъма K-Means в различен ред, постигайки средно силуетен коефициент 0.98 върху набор от данни от 661 потвърдени M&A сделки от 1975 до 2025 г. и разделяйки пространството на търсене на стратегии за придобиване на пет потвърдени клъстера, означени като High-Tech Targets, Growth-Stage Scale, Mature Portfolio, Young Startup Tuck-ins и Rapid High-Value. Второ, композитната метрика TechImpactScore интегрира патентни, финансови, времеви и стратегически признаци, използвайки тегла, извлечени от първата главна компонента на анализ на главните компоненти (PCA) върху признаците на придобиванията. Трето, сингулярният спектрален анализ (SSA) се прилага към времевия ред на M&A за първи път в тази литература; техниката декомпозира реда на значими трендове, цикли и сезонни компоненти и поддържа прогнозиране на бъдещи случвания. Четвърто, полуавтоматизирана многоагентна система, базирана на големи езикови модели (LLM), извлича нова информация от уеб, непрекъснато актуализирайки и разширявайки набора от данни на платформата при оперативни разходи под пет USD на месец с минимален брой валидации с човек в цикъла (human-in-the-loop). Накрая, всички предложени методи са валидирани и интегрирани в усъвършенствана версия на широко използваната уеб платформа EDA-DYV, поддържайки по-нататъшно изследване на знанието, извлечено от M&A сделките, и стратегическите прозрения, които то произвежда.

Практическа приложимост

Онлайн изследователска платформа EDA-DYV. Онлайн изследователската платформа EDA-DYV интегрира всеки аналитичен метод, разработен в този дисертационен труд, в единен уеб интерфейс с трислойна архитектура (C# .NET в бекенда с Entity Framework, и React с D3.js визуализации във фронтенда, SQL Server като база данни). Платформата показва клъстерен анализ, прогнозиране на вълни, анализ на преживяемостта, мрежови визуализации, анализ на времевата линия на M&A и полуавтоматизираната система за откриване и е внедрена на адрес <https://edadynamics.com>.

Публикации

Основните резултати на дисертационния труд са публикувани в осем научни труда, изброени в края на този автореферат. Те включват статии в списания и конференции, индексирани в Scopus, Web of Science и IEEE Xplore.

Структура и обем на дисертационния труд

Дисертационният труд има обем от 125 страници и се състои от въведение, пет глави, заключение, изложение на научните, научно-приложните и приложните приноси, раздел за ограниченията и бъдещата работа, и библиография от 120 източника. Той съдържа 21 фигури и 12 таблици. Номерацията на фигурите и таблиците в автореферата съответства на тези в дисертацията.

II. СЪДЪРЖАНИЕ НА ДИСЕРТАЦИОННИЯ ТРУД

ГЛАВА I. Аналитичен литературен обзор на динамиката на M&A, методите за клъстеризация и оценката на жизнеспособността в индустрията EDA

Автоматизираното проектиране (EDA) обхваща категорията софтуерни инструменти, използвани за проектиране на електронни системи като интегрални схеми (ИС) и печатни платки (PCB). Тя е фундаментален елемент във веригата на доставки на информационните и комуникационните технологии (ИКТ) и е еволюирала през последните 60 години заедно с експоненциално нарастващата сложност на полупроводниците. Действително, без напредъка в EDA проектирането на днешните процесори с техните милиарди транзистори би било немислимо. Индустрията EDA има мултидисциплинарен характер, съчетавайки аспекти на софтуерното инженерство, електроинженерството и компютърните науки, и е в сърцето на веригата на стойността на полупроводниковата индустрия, позволявайки много продукти и услуги.

Понастоящем индустрията EDA е тесен олигопол по отношение на масовите приложения за автоматизация на проектирането. Трите компании, доминиращи в тази област, са:

Synopsys, най-голямата EDA компания, има годишни приходи, надхвърлящи 5,8 милиарда долара (FY2023). Тя разполага с пълна гама от EDA инструменти, които покриват всеки етап от проектирането, включително синтез на ниво регистърен трансфер RTL (Design Compiler), разполагане и трасиране (IC Compiler), симулация (VCS) и IP. Synopsys е била най-активната в придобиванията сред основните EDA доставчици, осъществявайки повече от 80 придобивания през годините.

Cadence Design Systems, втората по големина EDA компания, генерира 4,1 милиарда долара приходи за FY2023. Virtuoso е водещият инструмент в категорията на персонализираното проектиране на ИС, Innovus води в цифровата имплементация, а Tempus в signoff верификацията. Компанията се разшири в проектирането и анализа на системи чрез придобивания, включително AWR Corporation (2011) за RF/микровълново проектиране и Numeca International (2021) за изчислителна флуидна динамика.

Siemens EDA (бивша Mentor Graphics) стана част от Siemens AG след придобиването ѝ за 4,5 милиарда долара през 2017 г. Бизнесът остава пазарен лидер в проектирането на PCB (Xpediton), IC верификацията (Questa) и проектирането за производство (Calibre). Интегрирането ѝ в портфолиото Siemens Xcelerator позволява нови приложения на цифров двойник и интелигентно производство.

В продължение на повече от десетилетие комбинираният пазарен дял на тримата водещи доставчици е близо 70%. Тази стойност беше идентифицирана рано в процеса на компилиране на базата с данни с транзакции и тя се прилага за трите възприети сегментации: по година на придобиване, по регион и по продуктова категория. Устойчивостта на тази концентрация във времето показва, че условията за самостоятелно оцеляване в EDA са не само трудни, но и структурно самоподсилващи се. Според Gao и Iyer траекторията на консолидация на индустрията EDA е подобна на тази на утвърдените технологични сектори, където предимствата на мащаба в R&D и взаимоотношенията с клиентите водят до самоподсилващи се бариери. Резултатите, представени тук, базирани на 661 транзакции за 50 години, дават малко основания за несъгласие.

Аналитичният литературен обзор, представен в тази глава, обобщава съществуващото знание за индустрията EDA в пет ключови области: (i) историята и пазарната структура на EDA; (ii) методите на машинно обучение и статистическите методи, приложими към анализа на M&A; (iii) бизнес моделите на EDA и рамките за оценка на жизнеспособността; (iv) динамиката на M&A и моделите на консолидация; и (v) приложенията на ИИ и големите езикови модели в технологичната индустрия.

Обзорът документира радикалната трансформация, която индустрията EDA е претърпяла през последните петдесет години. През 1980 г. индустрията представляваше съвкупност от доставчици на специфични за приложението инструменти, работещи на фрагментиран пазар. След десетилетия сливания и придобивания днес индустрията е близо до олигопол от трима играчи: Synopsys, Cadence Design Systems и Siemens EDA (бивша Mentor Graphics). По-голямата част от консолидацията се е случила чрез M&A и между 1998 и 2022 г. са идентифицирани четири вълни на сливания, всички корелиращи с цикли на полупроводниковите технологии.

Ключови находки от литературния обзор: Първо, съществува богатство от изследвания за M&A като цяло и за иновациите, движени от придобивания, във високотехнологичните индустрии. Нито едно от тях обаче не прилага съвременни подходи на машинно обучение към анализа на M&A в индустрията EDA; съществуващите трудове разчитат на традиционни статистически методи и ръчна класификация. Второ, голям обем работа по алгоритмите за клъстеризация и ансамбловите методи потвърждава предимствата на хибридните подходи. Нито един от тях не е проектиран за специфичните характеристики на данните за M&A транзакции: малки извадки, смесени типове признаци и стратегическата значимост на отдалечените транзакции. Трето, докато методи за прогнозиране като SSA са широко използвани в климатологията и икономиката, нито едно изследване не ги е използвало за предсказване на вълни на сливания и придобивания в технологичната индустрия. Четвърто, общата литература често е използвала техники за анализ на жизнеспособността за предсказване на корпоративен провал. Няма обаче публикувано изследване, фокусирано върху сектора EDA, където способността на дадена компания да остане конкурентоспособна с развитието на технологичните възли на полупроводниците е също толкова важна, колкото и нейното финансово здраве. Пето, техниките за мрежов анализ предоставят мощни инструменти за разбиране на структурните взаимоотношения в M&A мрежите. Доколкото е известно на автора, нито един индекс за пазарна концентрация или теоретико-графова метрика за централност не е прилаган към мрежата на придобиванията в EDA. Шесто, въпреки че големите езикови модели вече правят полуавтоматизираното извличане на информация от неструктурирани източници осъществимо, нито едно настоящо изследване не използва тази технология за откриване на данни за M&A в индустрията EDA. Накрая, основополагащо изследване върху бизнес моделите на EDA събра първоначален набор от данни за M&A и класификация на EDA компаниите, а по-ранна работа прегледа техники за формализация за оценка на бизнес модели в EDA. Нито един от тези проекти обаче не изгради прогностични модели, не използва анализ, движени от ИИ, и не направи данните достъпни чрез онлайн изследователска платформа. Съвързвайки нишките заедно, от обзора следват седем конкретни пропуски в текущото знание:

Пропуск 1 – Анализът на M&A модели, движен от ИИ, липсва в литературата за EDA. Съществуващите изследвания са описателни и не могат да извличат латентни стратегически модели от големи набори от данни за трансакции.

Пропуск 2 – Не е проектирана хибридна техника за клъстеризация за данни за M&A. Отделни алгоритми като гаусови смесени модели, DBSCAN и K-Means са използвани широко, но нито един хибриден подход не комбинира клъстеризация, базирана на плътност, и партиционна клъстеризация, за да обработи шума, въведен от стратегически значимите отдалечени придобивания.

Пропуск 3 – Не съществува валидирана композитна метрика за количествено определяне на технологичното въздействие на дадена M&A трансакция. Стойността на сделката, броят на патентите и възрастта на компанията се използват рутинно изолирано, но не са били комбинирани в единна валидирана композитна мярка.

Пропуск 4 – Не съществува модел за прогнозиране на вълни на сливания и придобивания за сектора EDA. Макар теорията за вълните на сливанията да е добре установена, нито едно изследване не е използвало методи за декомпозиция на времеви редове за предсказване на вълните на сливания в сектора EDA.

Пропуск 5 – Не съществува рамка за прогнозиране на жизнеспособността, калибрирана за EDA компании. Общата литература предоставя модели за прогнозиране на корпоративен провал, но нито един не разглежда специфичните за EDA детерминанти на жизнеспособността, свързани със съответствието с технологичния цикъл и вероятността от придобиване.

Пропуск 6 – Екосистемата на придобиванията в EDA не е изучавана като мрежа. Моделите на централност на придобиващите и динамиката на пазарната концентрация остават неизследвани.

Пропуск 7 – Не съществува автоматизиран инструмент за събиране на данни за M&A за изследователите в EDA. Събирането на данни е изцяло ръчно и отнема много време, и не е наличен полуавтоматизиран инструмент с помощта на ИИ.

Пропуск 8 – Не съществува интегрирана онлайн изследователска платформа, която да комбинира данни за M&A в EDA, клъстерен анализ, модели за прогнозиране, мрежова визуализация и оценка на жизнеспособността в единен ресурс.

Пропуск 9 – Методите, разработени за индустрията EDA, не са валидирани в съседни области, така че тяхната обобщаемост и преносимост остават непотвърдени.

ГЛАВА II. Методология на изследването и математически методи за анализ и прогнозиране на модели в M&A

Методи за клъстеризация и разпознаване на образи

Алгоритъм за клъстеризация K-Means

Алгоритъмът K-Means итеративно минимизира вътрешноклъстерната сума на квадратите (WCSS): за да раздели n наблюдения на K клъстера,

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^K \sum_{\vec{x} \in S_i} \|\vec{x} - \vec{\mu}_i\|^2 \quad \#1$$

където S_i представлява множеството от точки, разпределени към клъстер i , а μ_i , центроидът на клъстера S_i , се изчислява чрез вземане на средната стойност на всички точки в клъстера,

$$\vec{\mu}_i = \frac{1}{|S_i|} \cdot \sum_{\vec{x} \in S_i} \vec{x} \quad \#2$$

Инициализират се K центроида, всяка точка се присвоява към най-близкия центроид, центроидите се преизчисляват като средни на клъстерите и процесът се повтаря до постигане на сходимост. Реализацията използва трейнъра K-Means в ML.NET, който прилага стратегията за инициализация K-Means++, за да предотврати неоптимални локални минимума. Избор на оптимално K. Идеалният брой клъстери се намира чрез метода на лактя. Точката на „лактя“, в която маржиналното намаление на WCSS спада, се избира като идеалното K, когато WCSS се нанесе спрямо K за стойности в диапазона от 2 до 10. Методът на лактя определя, че K = 5 е идеалната стойност за набора от данни за M&A в EDA (n = 661).

Алгоритъм DBSCAN

В DBSCAN клъстерите се дефинират чрез локална плътност. Ако в епсилон-околността на точка p има поне MinPts точки, тогава тази точка се счита за ядрена точка.

$$|N_{\epsilon}(p)| \geq \text{MinPts} \#3$$

където епсилон-околността се дефинира като:

$$N_{\epsilon}(p) = \{q \in D: \text{dis}(p, q) \leq \epsilon\} \#4$$

За класифициране на точките се използват три категории: ядрени точки са тези, които удовлетворяват горния критерий за плътност; гранични точки са тези, които се намират в епсилон-околността на ядрена точка, но самите те не са ядрени; а шумови точки са тези, които не са нито ядрени, нито гранични. Съвързането на достижими по плътност ядрени точки създава клъстери. В случай че съществува верига от ядрени точки, в която всяка следваща двойка е на разстояние, по-малко от епсилон, тогава две ядрени точки, p и q, принадлежат към един и същ клъстер.

Методи за прогнозиране и анализ на преживяемостта

Сингулярен спектрален анализ

Етап 1: Вграждане. Едномерният времеви ред $X = (x_1, x_2, \dots, x_N)$ се преобразува в траекторна матрица с помощта на дължина на прозореца L:

$$X = (x_1 \ x_2 \ \dots \ x_K \ x_2 \ x_3 \ \dots \ x_{K+1} \ \vdots \ \vdots \ x_L \ x_{L+1} \ \dots \ x_N) \#6$$

където $N - L + 1 = K$. В съответствие с препоръката, че L трябва да бъде между N/3 и N/2, дължината на прозореца за тримесечния времеви ред за M&A в EDA е зададена на L = 12, което представлява три години тримесечни данни.

Стъпка 2: Сингулярно разлагане. SVD се използва за разлагане на траекторната матрица.

$$X = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T \#7$$

където U_i са левите сингулярни вектори, V_i са десните сингулярни вектори, а λ_i са собствените стойности (в намаляващ ред). Всяка тройка $(\sqrt{\lambda_i}, U_i \text{ и } V_i)$ се нарича собствена тройка. Стъпка 3: Формиране на групи. Компонентите на собствените тройки се категоризират според шум, осцилаторни цикли и тенденции. Обикновено тенденцията се улавя от първата собствена тройка, цикличните компоненти се улавят от следващите двойки, а шумът се представя от останалите собствени тройки. След това реализацията следва тенденцията и двата основни периодични компонента в данните, като задава броя на запазените собствени тройки на 3 (зададено чрез параметъра NUM_COMPONENTS = 3).

Оценител на преживяемост на Каплан–Майер

Оценителят на Каплан–Майер е непараметричен статистически оценител на функцията на преживяемост от данни за продължителността на живота. В контекста на компаниите от EDA събитието на преживяемост се дефинира като независимост от поглъщане от други компании (да не бъдат придобити или ликвидирани).

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \quad \#8$$

d_i е броят на събитията (придобивания) в момент t_i , а n_i е броят на бизнесите под риск (които все още работят независимо) непосредствено преди момент t_i . Кривата на Каплан–Майер визуализира вероятността за преживяемост на компания от EDA на определена възраст.

Модел на пропорционалните рискове на Кокс

Моделът на пропорционалните рискове на Кокс е развитие на оценителя на Каплан–Майер и включва допълнителни ковариати, за да отчете факторите, които могат да повлияят на резултатите за времето до настъпване на събитие. Функцията на риска на бизнес с вектор от признаци X в момент t е следната.

$$h(\vec{X}) = h_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \quad \#9$$

където коефициентите β_i количествено определят влиянието на всяка ковариата върху интензитета на риска, а базовата функция на риска, $h_0(t)$, е една и съща за всички компании. Докато положителна стойност на β показва, че по-високите стойности на съответната ковариата увеличават риска (риск от придобиване), отрицателна стойност на β показва защитен ефект. Моделът за прогнозиране на жизнеспособността в EDA използва редица ковариати, включително възраст на компанията, TechImpactScore, позиция по централност в мрежата (от мрежовия анализ) и принадлежност към клъстер (от обратното хибридно клъстериране (Reverse Hybrid Clustering)). Моделът на Кокс се валидира с помощта на конкордансния индекс (C-индекс), който оценява доколко добре моделът подрежда двойки бизнеси въз основа на очаквания им риск. Перфектна конкордантност се обозначава с C-индекс 1.0, докато случайно предсказание се обозначава с C-индекс 0.5. Brier Score: показва степента на калибриране на предсказаните вероятности за преживяемост; по-ниските стойности показват по-добро калибриране. Базовите интензитети на риска се извличат от емпиричните специфични за клъстера темпове на придобиване, изчислени от историческия набор от данни за M&A в EDA. След това тези темпове се прилагат върху петгодишен прогнозен хоризонт.

Методи за мрежов анализ и статистическо валидиране

Мрежов анализ

Централността по степен изчислява броя на преките връзки на даден възел, нормализиран спрямо максималния брой връзки, които могат да съществуват.

$$C_D(v) = \frac{\deg(v)}{n-1} \quad \#10$$

където n е общият брой върхове, а $\deg(v)$ е броят на ребрата, инцидентни на връх v . Серийните придобиващи се идентифицират чрез висока централност по степен в мрежата на придобиванията.

Честотата, с която даден възел се намира на най-късия път между други възли, се измерва чрез неговата централност по посредничество.

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad \#11$$

където $\sigma_{st}(v)$ е броят на пътищата, които преминават през v , а σ_{st} е общият брой на най-късите пътища от възел s до възел t . „Мостовите“ компании, които свързват различни сегменти на мрежата на придобиванията, се идентифицират чрез висока централност по посредничество. Средното разстояние между даден възел и всеки друг възел се измерва чрез централността по близост.

$$C_C(v) = \frac{n-1}{\sum_{u \neq v} d(v, u)} \quad \#12$$

където разстоянието по най-къс път между v и u се обозначава с $d(v, u)$. Бизнесите, които са добре позиционирани да достигат бързо до всяка част от мрежата, се идентифицират чрез висока централност по близост. Идеята, че връзките към силно централни възли са по-ценни от връзките към периферни възли, се улавя от собственовекторната централност, която е разширение на централността по степен, претеглящо връзките според централността на съседите.

Индекс на Херфиндал–Хиршман

Пазарната концентрация на пазара на придобивания в EDA се измерва чрез индекса на Херфиндал–Хиршман (HHI).

$$HHI = \sum_{i=1}^N s_i^2 \quad \#13$$

където s_i представлява пазарния дял на фирма i като процент от всички придобивания. HHI е между 1 (монопол) и $1/N$ (съвършена конкуренция). При наблюдение на пазарите в САЩ, тези с HHI под 0.15 се считат за неконцентрирани, тези между 0.15 и 0.25 за умерено концентрирани, а тези над 0.25 за силно концентрирани, съгласно насоките на Министерството на правосъдието

Метрики за валидиране на клъстери

Кохезията и разделянето на клъстера за всяка точка от данните се измерват чрез силуетния коефициент

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad \#14$$

където $b(i)$ е средното разстояние до най-близкия клъстер (средно разстояние до всички точки в най-близкия съседен клъстер), а $a(i)$ е средното вътрешноклъстерно разстояние (средно разстояние до всички останали точки в същия клъстер). Силна клъстерна структура се обозначава със стойности над 0.7, които варират от -1 до 1. Индексът на Дейвис–Боулдин изчислява отношението на междуклъстерното разделяне към вътрешноклъстерното разсейване.

$$DB = \frac{1}{K} \sum_{i=1}^K \frac{(\sigma_i + \sigma_j)}{d(c_i, c_j)} \quad \#15$$

където $d(c_i, c_j)$ е разстоянието между центроидите, а σ_i е средното разстояние между точките в клъстер i и техния центроид c_i . По-ниските стойности показват по-добро разделяне на клъстерите. Критерият на отношението на дисперсиите, или индексът на Калински–Харабаш, изчислява отношението на вътрешноклъстерната дисперсия към междуклъстерната дисперсия.

$$CH = \frac{\frac{SS_B}{K-1}}{\frac{SS_W}{n-K}} \quad \#16$$

където K е броят на клъстерите, n е общият брой точки от данни, а SS_B и SS_W са съответно междуклъстерната и вътрешноклъстерната сума на квадратите. По-високите стойности показват по-компактни и по-добре разделени клъстери.

Методи за статистическа валидиране

За да се определи дали съществуват значими разлики между идентифицираните клъстери по важните променливи (Lifetime, TechImpactScore, DaysSinceEarliest), се използва дисперсионен анализ (ANOVA). Тъй като $\mu_1 = \mu_2 = \dots = \mu_K$, нулевата хипотеза H_0 гласи, че всички средни на клъстерите са равни. $= \mu_K$. Тестовата статистика е следната.

$$F = \frac{MS_{between}}{MS_{within}} = \frac{\frac{SS_B}{K-1}}{\frac{SS_W}{n-K}} \quad \#17$$

където междугруповата сума на квадратите е:

$$SS_B = \sum_{k=1}^K n_k (\bar{x}_k - \bar{x})^2 \quad \#18$$

където $\bar{x}_k$ е средната на клъстера, $\bar{x}$ е общата средна, а n_k е броят на наблюденията в клъстер k . Сумата на квадратите в рамките на групата е:

$$SS_W = \sum_{k=1}^K \sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)^2 \quad \#19$$

При нулевата хипотеза F-статистиката съответства на F-разпределение със степени на свобода ($K - 1$, $n - K$). Прагът за определяне на статистическа значимост е $\alpha = 0.05$. Ета-квадрат (η^2) се използва за докладване на размерите на ефекта и се дефинира както следва.

$$\eta^2 = \frac{SS_B}{SS_B + SS_W} \quad \#20$$

В тази работа размерът на ефекта се докладва чрез ета-квадрат (η^2). Това, което той дава, е делът на общата дисперсия, който може да се отдаде на принадлежността към клъстер, с други думи частта, която идва от разликите между клъстерите, а не от разликите вътре във всеки клъстер. Cohen предлага три референтни стойности за интерпретация, като η^2 от около 0.01 означава малък ефект, 0.06 - среден, а 0.14 - голям; макар че трябва да се отбележи, че това са насоки, а не фиксирани граници. Когато данните не са нормални, допускането за нормалност на ANOVA не е безопасно и вместо него се използва тестът на Крускал–Уолис, тъй като той не изисква това допускане. Начинът, по който работи този тест, е чрез замяна на първоначалните наблюдения с техните рангове (изчислени съвместно за всички групи) и след това проверка дали тези рангове са разпределени различно между групите. Всичко това се обобщава в едно-единствено число, H-статистиката, изчислена чрез:

$$H = \frac{12}{n(n+1)} \sum_{k=1}^K \frac{R_k^2}{n_k} - 3(n+1) \quad \#21$$

K във формулата по-горе представлява броят на групите. Броят на наблюденията, принадлежащи към определена група k , се дава от n_k , докато R_k съдържа ранговата сума в рамките на тази група. Цялата извадка, за всички групи, се преброява от n . С $K - 1$ степени на свобода, H приблизително следва хи-квадрат разпределение при нулевата хипотеза, че всички групи имат еднакви разпределения. На свързаните стойности се присвояват средни рангове, за да се обработят свързаните наблюдения. За да се определи дали наблюдаваното качество на клъстеризацията (както е обозначено от силуетния коефициент) е по-високо от това, което би се предсказало по случайност, се използва пермутационен тест. За всяка от B пермутациите ($B = 1,000$) се създават случайни присвоявания към клъстери и се изчисляват съответните силуетни коефициенти. Така се определя p -стойността:

$$p = \frac{\#\{s_{random} \geq s_{observed}\} + 1}{B + 1} \quad \#22$$

където s_{random} представлява силуетния коефициент за всяка пермутация, а s_{obs} представлява наблюдавания силуетен коефициент. За да се предотврати нулева p -стойност и да се вземе предвид самата наблюдавана статистика, числителят и знаменателят се коригират с +1. Като оценка на валидността на клъстеризацията, независима от разпределението, този тест допълва непараметричния метод на Крускал–Уолис и параметричния метод ANOVA. Бутстрап произвадка със

100 итерации се използва за оценка на стабилността на резултатите от клъстеризацията. Всяка итерация включва вземане на случайна извадка с размер n от набора от данни, прилагане на алгоритъма за обратно хибридно клъстериране (Reverse Hybrid Clustering) със същите параметри и използване на коригирания индекс на Ранд (ARI) за сравняване на крайните присвоявания към клъстери с първоначалните.

$$ARI = \frac{RI - E[RI]}{ma(RI) - E[RI]} \quad \#23$$

където $E[RI]$ е очакваният индекс на Ранд при случайно присвояване, а RI е индексът на Ранд (делът на двойките точки, които са или в един и същ клъстер и в двете разбивания, или в различни клъстери и в двете). При ARI -1 показва антикорелация, 0 представлява съгласие на ниво случайност, а 1 представлява перфектно съгласие. В подкрепа на тези твърдения беше проверена консистентността на ARI коефициента между производките, за да се удостовери, че всяка наблюдавана клъстерна структура не е специфична за извадката. Следващата стъпка беше да се идентифицира кой от отделните признаци допринася най-много за TechImpactScore. Беше извършена аблация на признаци, като всеки от четирите признака (Patent Score, Deal Value, Age Score, Acquirer Frequency) се премахваше последователно и силуетният коефициент се преизчисляваше, за да се види как се променят получените клъстери. Големите спадове в коефициента показват важни признаци, а малките спадове показват признаци, които не са важни.

ГЛАВА III. Обратно хибридно клъстериране, композитни метрики, предиктивни модели и многоагентна система за откриване на M&A

Този дисертационен труд предлага нов алгоритъм за обратно хибридно клъстериране (Reverse Hybrid Clustering) за данни от M&A сделки, тристепенен ансамблов подход за клъстеризация, който интегрира DBSCAN и K-Means по нов последователен начин. Точките от данните се клъстеризират в три етапа: клъстеризация по плътност с DBSCAN, прецизиране с K-Means и реинтеграция на шумовите точки. В този раздел се обсъждат параметрите, предварителната обработка на данните и техниките за оптимизация, използвани в предложения подход за обратно хибридно клъстериране. Повечето хибридни техники за клъстеризация, използвани понастоящем в литературата, прилагат права конфигурация, при която DBSCAN се използва за локално прецизиране по плътност, след като K-Means или йерархичната клъстеризация се използват за установяване на глобална структура. Настоящото изследване обръща тази последователност въз основа на специфична за областта находка: M&A активността в сектора на EDA показва изблици, основани на естествена плътност, които съвпадат с вълните на пазарна консолидация, за разлика от равномерно разпределените времеви модели. Алгоритъмът първо използва DBSCAN, за да улови тези модели на естествена плътност, преди да наложи предварително определен брой клъстери.

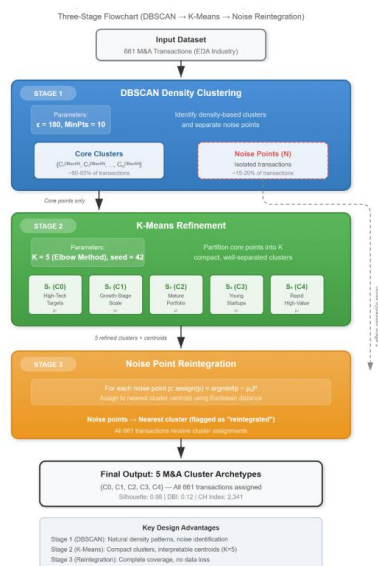
Аблационно изследване, представено по-нататък в тази глава, осигурява емпирична подкрепа за това проектно решение, като показва, че върху един и същ набор от данни обратната последователност постига силуетен коефициент 0.98, в сравнение с 0.72 за правия хибриден подход.

Всяка M&A сделка се представя чрез триизмерен вектор от признаци

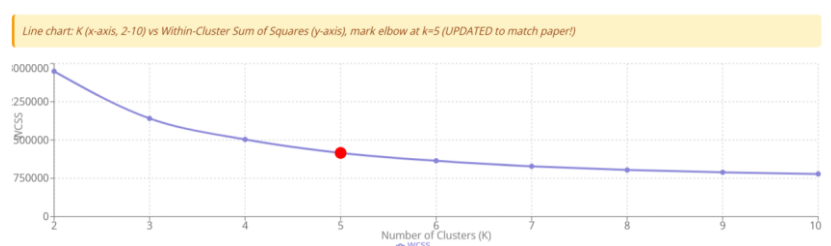
$$\vec{x}_i = (DaysSinceEarliest_i, Lifetime_i, TechImpactScore_i) \quad \#25$$

Lifetime: Времетраене на съществуване на целевата компания в години към момента на придобиването; DaysSinceEarliest: Положение на придобиването във времето спрямо най-ранната сделка в набора от данни; TechImpactScore: Композитен резултат, представящ въздействието на технологията върху придобитата компания. Всички стойности на признаците се нормализират в диапазона $[0, 10]$ преди клъстеризацията с Min-Max скалер. Първо се разработват спецификациите за реализация на клъстеризацията по плътност с DBSCAN (Етап 1). Този метод за клъстеризация различава изолираните сделки от регионите с висока плътност на придобивна активност в целия набор от 661 M&A сделки. За регионален растеж на обединените клъстери чрез стандартен DBSCAN

се прилагат две специфични за областта адаптации: времево претегляне и калибриране на MinPts. За времево претегляне измерението, представящо DaysSinceEarliest, се претегля, а epsilon се задава на 180, което е приблизително шест месеца във времево пространство и средната продължителност на изблик на M&A консолидация, по време на който няколко придобивания в рамките на конкретна индустрия се консолидират за относително кратък период от време. За калибрирането на MinPts се избира по-голяма стойност 10, за разлика от евристиката $\ln(N)$, изведена в, която дава $\ln(661) \approx 6.5$. MinPts беше зададен на 10, за да се обединят в единичен клъстер сделки, които отразяват стабилни и значими модели на консолидация, за разлика от сделки, които просто настъпват близо във времето. Такова увеличение на MinPts е необходимо, за да се избегнат фалшиви открития и за по-ясно идентифициране на основополагащите вълни на сливания и придобивания в данните. Множество клъстери, основани на плътност, и шумово множество N , съдържащо изолирани сделки, които не отговарят на критерия за плътност, са резултатите от Етап 1. Етап 1 на набора от данни за EDA M&A обикновено намира от три до четири плътни региона, където от 15 до 20 процента от сделките се категоризират като шум. Тристепенният поток на алгоритъма е обобщен графично на Фигура 1, която показва пътя на данните от суровите сделки през клъстеризацията по плътност с DBSCAN, прецизирането с K-Means върху ядрените точки и реинтеграцията на шума до окончателно присвояване $\{C_1^{DBSCAN}, C_2^{DBSCAN}, \dots, C_m^{DBSCAN}\}$



Фигура 1 – Алгоритъм за обратно хибридно клъстериране – тристепенна блок-схема.



Фигура 2 – Графика по метода на лакътя за избор на оптимално K.

$$\{C_1^{DBSCAN}, C_2^{DBSCAN}, \dots, C_m^{DBSCAN}, N\} = DBSCAN(\vec{X}, \epsilon, MinPts) \quad \#26$$

където C_j ($j = 1, 2, \dots, m$) е j-тият клъстер, основан на плътност, а N е множеството от шумови точки. Използвайки два параметъра, epsilon и MinPts (зададени съответно на 180 и 10), методът DBSCAN е в състояние да идентифицира естествения модел на плътност в данните и да различи шума (отклоненията) от типичните изблици на консолидация в M&A активността. На втория етап методът K-Means се изпълнява върху обединението на ядрените точки на всички клъстери, намерени на първия етап. В резултат от изпълнението на процеса K-Means ядрените точки се клъстеризират в K

клъстера, които са относително компактни, като характеристиките на всеки клъстер се правят особено ясни чрез неговите центроиди:

$$\{S_1, S_2, \dots, S_K\} = \left(\bigcup_{j=1}^m C_j^{DBSCAN}, K \right) \quad \#27$$

Методът на лакътя, приложен върху ядрените точки, дава $K = 5$ като оптимален брой клъстери за набора от данни за EDA M&A. Етап 3 реинтегрира всяка шумова точка $p \in N$ към най-близкия центроид на прецизиран клъстер, използвайки евклидово разстояние, при условие че разстоянието е под праг на сходство τ :

$$(p) = \underset{k \in \{1, \dots, K\}}{\operatorname{argmin}} \|\vec{p} - \vec{\mu}_k\|^2$$

$$\text{if } \|\vec{p} - \vec{\mu}_k\| \leq \tau \quad \#28$$

Epsilon	MinPts	K	Силуетен коефициент	DBI	СН индекс
120	8	4	0.91	0.21	1,847
150	10	5	0.95	0.15	2,156
180*	10*	5*	0.98*	0.12*	2,341*
180	12	5	0.96	0.14	2,198
210	10	5	0.94	0.18	2,087
240	10	6	0.89	0.22	1,923

Таблица 2 – Резултати от мрежово търсене на параметри за обратно хибридно клъстериране

TechImpactScore: Композитна метрика за технологична оценка на M&A

Технологичното въздействие на M&A върху фирмите и икономиката все още не е количествено оценено в литературата. Макар да съществува обширна литература, която оценява M&A дейностите чрез стойност на сделката и приходни множители, в тази статия се въвежда нова композитна метрика, наречена TechImpactScore, за измерване на технологичното въздействие на сделките. Стойността на сделката и приходните множители са двете най-широко използвани метрики за оценка на резултатите от сливанията и придобиванията във финансовата общност. Факторите на стойността, които влияят върху стратегическите и тактическите цели в зависещите от технологии индустрии обаче, не винаги се отразяват в тези две метрики, които може да не разкриват напълно значимостта на придобитите технологии. Конкретна празнина в прегледа на литературата се запълва от метриката: настоящите аналитични рамки за M&A третират всички придобивания като еднородни събития само с финансовата стойност като определяща характеристика. Стратегическата и технологичната значимост на придобиванията в сектора на EDA обаче варира значително. Например стратегическите последици от придобиване за 50 милиона долара на предприятие с 200 патента и 15 години експертиза в областта се различават съществено от тези на придобиване за 50 милиона долара на стартъп, който е само на две години и няма патентно портфолио. Тези качествени вариации се улавят от TechImpactScore, количествена композитна метрика. TechImpactScore се дефинира като претеглена сума от четири нормализирани компонента: патентния резултат, резултата по стойност на сделката, резултата по възраст на компанията и резултата по честота на придобивания:

$$TechImpactScore = 0.35 \cdot P_n + 0.30 \cdot D_n + 0.20 \cdot A_n + 0.15 \cdot F_n \quad \#29$$

където P_n е нормализираният патентен резултат (данни от API на USPTO PatentsView), D_n е нормализираният резултат по стойност на сделката, A_n е нормализираният резултат по възраст на компанията, а F_n е нормализираният резултат по честота на придобивания. Всеки компонент се нормализира до скала $[0, 10]$ чрез Min-Max нормализация:

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \times 10 \quad \#30$$

Теглата (0.35, 0.30, 0.20, 0.15) са получени от PCA върху 661 M&A сделки. Натоварванията на първата главна компонента показват приноса на всяко измерение към общата дисперсия: Патентният резултат има натоварване 0.38 (нормализирано тегло 0.35), Стойността на сделката 0.32 (нормализирано тегло 0.30), Възрастта на компанията 0.18 (нормализирано тегло 0.20, коригирано нагоре за релевантност към областта) и Честотата на придобивания 0.12 (нормализирано тегло 0.15, също коригирано нагоре

за релевантност към областта). За да се гарантира, че всеки компонент има значим минимален принос, изведените чрез PCA тегла бяха незначително модифицирани въз основа на експертиза в областта, в съответствие с методологията, предложена от Nardo et al. за изграждане на композитни индикатори. Валидността на TechImpactScore беше проверена чрез сравняването му с независими мерки за стойност на сделката (т.е. покупната цена или стойността на транзакцията на сделката). Използваните данни бяха предоставени за всяка сделка, за която беше наличен TechImpactScore. Беше установена корелация на Пърсън $r = 0.71$ ($p < 0.001$). Макар че корелация, която е по-малка от перфектна, обикновено се счита за умерена, находката е точно това, което би се очаквало, за да укаже конвергентна валидност на TechImpactScore като мярка за едно конкретно измерение на стойността на придобиването, отделно от стойността на сделката, но корелиращо с нея. Това подсказва, че резултатът улавя технологични и стратегически фактори в допълнение към стойността на сделката.

Откриване на пет архетипа на M&A клъстери

С помощта на алгоритъма за обратно хибридно клъстериране, приложен върху 661 EDA M&A сделки, бяха разкрити пет отделни клъстерни модела за стратегическо придобивно поведение, или клъстерни архетипи, които представят уникални стратегии за придобиване. Тези архетипи съставляват нова емпирична таксономия на M&A поведението в рамките на технологично интензивните индустрии.

Клъстер	Наименование	Размер (%)	Среден жизнен цикъл (години)	Среден брой патенти	Среден TechImpactScore	Стратегическа интерпретация
C0	Високотехнологични цели	15.1%	11.2	18.6	7.8	Зрели технологични компании, придобити заради силни патентни портфейли (IP)
C1	Масщаб във фаза на растеж	35.9%	9.8	8.2	5.4	Най-голям клъстер; компании в средна фаза, придобити заради пазарен мащаб
C2	Зрял портфейл	18.7%	18.6	5.4	4.1	Отдавна установени компании, придобити заради клиентска база и пазарна позиция
C3	Присъединяване на млади стартъпи	20.4%	4.2	3.1	3.6	Стартъпи в ранна фаза, придобити основно заради таланти и нововъзникващи технологии
C4	Бързи високостойности	9.9%	<2.0	12.8	8.9	Много млади компании с изключителни стойности на сделките (средно \$249M); стратегически придобивания с премия

Таблица 3 – Пет архетипа на M&A клъстери, идентифицирани чрез обратно хибридно клъстериране.

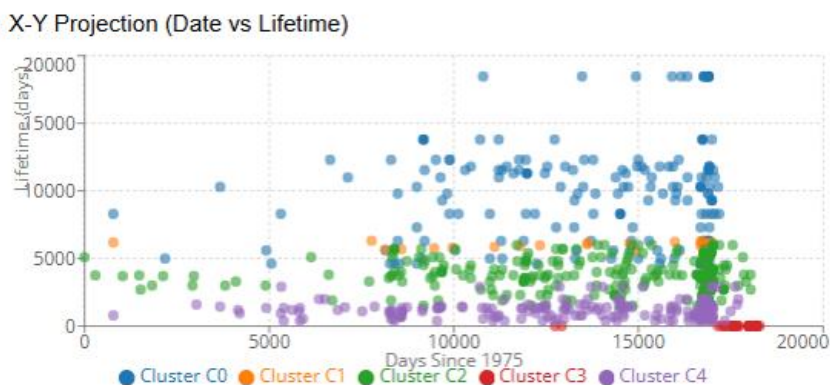
Клъстер C0 (High-Tech Targets) се използва за описание на придобиването на високотехнологични, обикновено патентно интензивни компании. Тези компании имат бизнес времетраене от 10 години или повече и притежават огромно количество IP и ноу-хау. Разглеждайки конкретно сектора на EDA, придобиванията C0 често включват IP, DFM или специализирани верификационни фирми, чиито технологични портфолия са трудни за вътрешно възпроизвеждане. Придобиването на IP е основната стратегическа мотивация, както свидетелства средната стойност от 18.6 патента на цел.

Повече от една трета от всички сделки бяха насочени към компании в Клъстер C1 (Growth-Stage Scale), най-големия клъстер. Това са компании с умерени количества технологични активи във фаза на растеж, на приблизително 10 години. Повечето покупки в този клъстер бяха на конкурентни компании, както и на доставчици на допълващи продукти, за да се увеличи обхватът на покритие на проектния поток на придобивания. Стратегическата движеща сила зад това е пазарната консолидация и мащаба. Сделките C1 формират ядрото на модела на консолидация в индустрията.

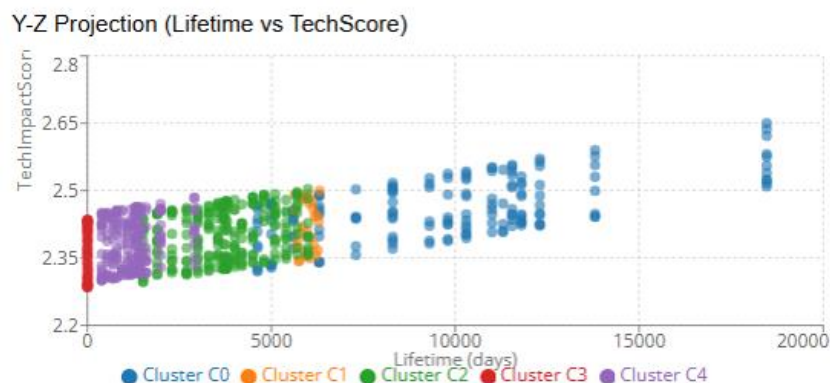
Придобиванията на утвърдени компании с по-малък брой патенти (средно на 18.6 години) се категоризират под Клъстер C2 (Mature Portfolio). Тези видове придобивания обикновено се извършват за завършване на технологична платформа, за разширяване в географски пазар и/или за придобиване на клиентска база. Когато конкретно се отнася до сектора на EDA, C2 обикновено включва покупката на по-малка EDA компания с установена клиентска база, обслужваща конкретна пазарна ниша.

Компаниите в Клъстер C3, Young Startup Tuck-ins (средна възраст: 4.2 години), са млади стартъпи с иновативна технология и способни инженерни екипи. Те са пример за „acqui-hire“ и съответстват на мотивацията за придобиване на таланти, описана в литературата. Стартъпите са в ранен етап от жизнените си цикли и имат ниска патентна производителност. Затова средният TechImpactScore е относително нисък - 3.6.

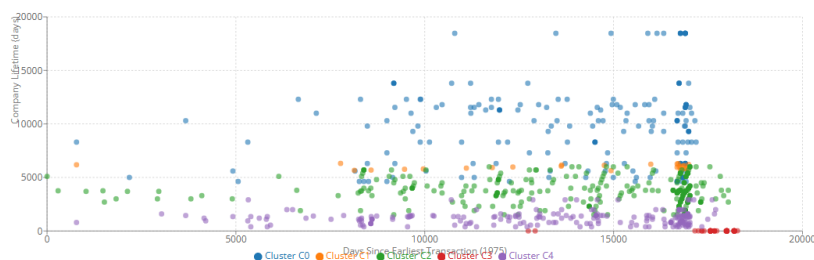
Клъстер C4 (Rapid High-Value), най-открояващият се архетип, е съставен от изключително млади предприятия със средна продължителност на съществуване под две години, които достигат изключително високи придобивни премии (средна стойност на сделката 249 милиона долара). Тези сделки са примери за стратегически залози върху преобразуващи технологии, при които купувачът плаща повече, за да получи потенциално революционна технология преди конкурентите. Придобиванията C4 в сектора на EDA често включват нова полупроводникова IP или инструменти за проектиране с поддръжка на AI/ML. Високото патентно качество, а не количество, се отразява във високия TechImpactScore (8.9). Пространственото разделяне на петте архетипа в пространството на признаците е визуализирано на Фигура 3а и Фигура 3б, където всяка сделка е нанесена в триизмерното пространство, дефинирано от DaysSinceEarliest, Lifetime и TechImpactScore, и е оцветена според клъстерния етикет.



Фигура 3а – Пет архетипа на M&A клъстери



Фигура 3b – Пет архетипа на М&А клъстери – триизмерна точкова диаграма



Фигура 4 – Времево разпределение на М&А сделките по клъстер.

Конфигурация	Силуетен коефициент	DBI	СН индекс	Влош. (%)
Всички признаци - базова линия	0.98	0.12	2341	-
Без патентни данни	0.82	0.28	1645	-16.3%
Без стойност на сделката	0.87	0.23	1812	-11.2%
Без възраст на компанията	0.91	0.19	1987	-7.1%
Без TechImpactScore	0.64	0.45	1023	-34.7%

Таблица 4 – Резултати от аблационното изследване на признаците

Алгоритъм	Силуетен коефициент	DBI	СН индекс	Обработка на шума	Интерпр. на клъстерите
K-Means (K=5)	0.72	0.34	1567	Няма	Висока
DBSCAN (epsilon=180, MinPts=10)	0.85	0.21	1823	Да	Средна
Право хибридно (K-Means → DBSCAN)	0.72	0.31	1678	Частична	Средна
GMM (K=5)	0.68	0.39	1456	Няма	Ниска
Обратно хибридно (DBSCAN → K-Means)	0.98	0.12	2341	Да	Висока

Таблица 5 – Сравнение на алгоритмите за клъстеризация

Прогнозиране на вълни чрез SSA: Адаптация към времеви редове на М&А

В този раздел се описва как сингулярният спектрален анализ е модифициран за откриване и прогнозиране на М&А вълни. Това е приложение на SSA към времеви редове на корпоративни сделки, което не е документирано в литературата преди в сектора на EDA. SSA е приложен обширно към финансови и икономически времеви редове; за да се справи с нерегулярните времеви модели, типични за придобивната активност, и с дискретния, основан на брой характер на данните за сделките, неговото приложение към данните за М&А активност налага конкретни модификации. Броят на сделките във всяко календарно тримесечие от Q1 1975 до Q4 2025 се преброява, за да се преобразуват суровите данни за М&А сделки в тримесечен времеви ред. Това води до времеви ред с

$N = 204$ наблюдения на тримесечие. Най-добрата времева резолюция за улавяне на моделите на M&A вълните е тримесечната агрегация, а не годишната (която би загубила вътрешногодишната динамика) или месечната (която би довела до прекомерни периоди с нулев брой). Три години тримесечни данни се представят чрез прозореца на вграждане, който е зададен на $L = 12$ тримесечия. Освен че е достатъчно голям, за да улови типичната продължителност от 3-5 години на M&A консолидационните вълни, наблюдавани в индустрията на EDA, той също така е в съответствие с общото указание, че L трябва да попада между $N/3$ и $N/2$. Поради разреденост на ранните сделки в набора от данни, по-дълъг прозорец ($L = 20$) въведе гранични ефекти, докато по-къс прозорец ($L = 8$) не беше в състояние да различи първичните и вторичните циклични компоненти. Компонентите на сигнала, които се запазват, и тези, които се отхвърлят като шум, зависят от броя на запазените собствени тройки. Въз основа на емпиричен анализ на спектъра от собствени стойности бяха запазени първите три собствени тройки: първата представя дългосрочната възходяща тенденция в M&A активността, втората представя първичния цикличен модел с период около 8-10 години, а третата представя вторична осцилация с период около 4-5 години. Заедно тези три компонента обясняват по-голямата част от структурата на дисперсията и премахват високочестотния шум, който би намалил точността на прогнозата. Рекурентният алгоритъм за прогнозиране проектира реконструирания сигнал 20 тримесечия (5 години) в бъдещето, произвеждайки прогнози от 2026 г. до 2030 г.. Доверителните интервали се оценяват чрез експоненциално изглаждане на Холт–Уинтърс, с параметрите $\alpha = 0.3$ (ниво) и $\beta = 0.1$ (тенденция). Използвайки прогнозната стойност ± 1.96 стандартни отклонения на остатъците между SSA реконструкцията и оригиналните данни, се изчисляват 95-процентовите интервали. Първа стъпка за валидиране на SSA декомпозицията на времевия ред на EDA M&A е да се провери дали цикличните компоненти, възстановени от алгоритъма, съответстват на четирите исторически M&A вълни, които са документирани независимо в литературата за EDA. Таблица 6 отчита за всяка от четирите вълни периода, през който вълната е активна, пиковата година, идентифицирана от SSA реконструкцията, пиковия тримесечен брой сделки при този пик и доминиращия индустриален двигател, свързан с вълната. Възстановяването на всичките четири вълни единствено от данните, с пикове, които съответстват на историческия запис в рамките на едно тримесечие, показва, че цикличните компоненти, извлечени чрез SSA, представят истинска пазарна динамика, а не статистически артефакти на декомпозицията.

Вълна	Период	Пикова година	Пиков тримес. брой	Двигател
Вълна 1	1998-2001	2000	28	Консолидация след дотком бума
Вълна 2	2006-2008	2007	22	Сложност на SoC преди финансовата криза
Вълна 3	2015-2018	2017	25	Възприемане на ИИ/облачни технологии
Вълна 4	2020-2022	2021	24	Недостиг на чипове заради Covid, дистанционно EDA

Таблица 6 – Исторически M&A вълни, открити чрез SSA

Анализ на преживяемостта за прогнозиране на жизнеспособността на компаниите

Версия на модела на пропорционалните рискове на Кокс се използва в рамката за анализ на преживяемостта за прогнозиране на жизнеспособността на EDA компаниите, където „събитие“ е придобиване или прекратяване, а „преживяемост“ се дефинира като „продължаване на независима дейност“. Интеграцията на разнородни аналитични признаци от анализите на клъстеризация, мрежа и технологии, създадени в този дисертационен труд, прави това приложение уникално. Чрез комбиниране на резултатите от няколко аналитични модула моделът на Кокс сглобява петизмерен вектор от признаци за всяка компания.

Признак 1 – Принадлежност към клъстер (категориален). Присвояването на клъстер от обратното хибридно клъстериране (C0-C4) улавя стратегическия архетип на компанията и свързания с него профил на придобивен риск.

Признак 2 – TechImpactScore (непрекъснат). Тази композитна метрика за технологична оценка съобщава на потенциалните купувачи значимостта на технологията на компанията.

Признак 3 – Позиция по мрежова централност (непрекъснат). Централността по посредничество на компанията в придобивната мрежа изразява нейната структурна значимост в M&A екосистемата.

Признак 4 – Възраст на компанията (непрекъснат). Броят на годините от основаването на компанията е ключов индикатор за придобивен риск в литературата за M&A.

Признак 5 – Категоризация по технологичен сектор. Индустриалният сегмент (например IP, DFM, симулация или верификация) влияе върху изложеността на компанията на конкуренти.

Това многоизмерно пространство на признаците е подобрение спрямо предишните приложения на анализ на преживяемостта в корпоративните финанси, които обикновено използват само финансови променливи. Моделът интегрира резултатите от технологичната оценка, мрежовата позиция и клъстерната принадлежност, за да отчете стратегически аспекти на финансовата жизнеспособност, които не са обхванати от традиционните индикатори (като възвръщаемост на продажбите и възвръщаемост на инвестицията).

Мрежов анализ на връзките между придобиващите

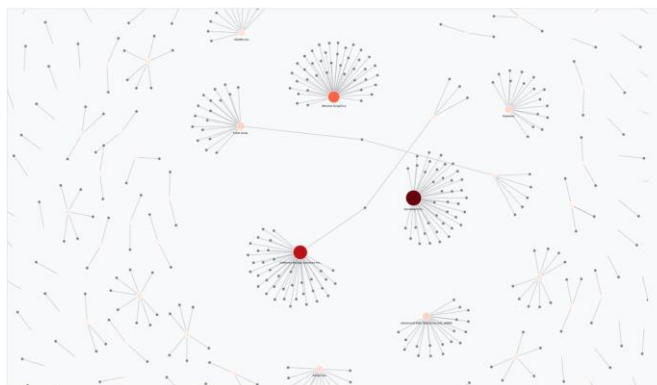
Както вече беше споменато, в модула за мрежов анализ се използват елементи на теорията на графите, за да се анализират структурните характеристики на придобивната екосистема на EDA. Тази цел се изпълнява от класа „NetworkAnalysisService.cs“. Ориентиран граф може да бъде конструиран по следния начин: Фирмите се представят чрез върхове, а придобивните връзки между фирмите се представят чрез ориентирани ребра, сочещи от придобиващата фирма към придобитата фирма. Придобивната мрежа е изградена от 661 M&A сделки, давайки граф с приблизително 600 уникални възела и 661 ребра. Мрежата е разреждана, тъй като повечето компании предприемат само 1-2 сделки (плътност < 0.01). Доминиращите доставчици на EDA са обкръжени от съзвездие от „хъб“ възли (сериини придобиващи) с висока изходяща степен в звездообразна конфигурация.

Ранг	Компания	Придобивания	Централност по степен	Централност по посредничество
1	Synopsys	80+	0.132	0.287
2	Cadence	50+	0.089	0.198
3	Mentor Graphics / Siemens EDA	40+	0.067	0.156
4	Intel	15+	0.028	0.089
5	Keysight Technologies	12+	0.021	0.067

Таблица 7 – Топ 5 придобиващи по централност по степен

Synopsys е най-активно придобиващият доставчик на EDA и така контролира мрежата по всички мерки за централност. Високата му централност по посредничество (0.287) означава, че Synopsys действа като ключов свързващ мост между различните пазарни сегменти на EDA, с разнообразно придобивно портфолио, което включва IP, DFM, симулация и верификация. За да се проследи еволюцията на пазарната концентрация в придобиванията на EDA, индексът на Херфиндал–Хиршман (HHI) се изчислява в рамките на времеви прозорци. В началото на 90-те години придобивният пазар на EDA започна като неконцентриран пазар ($HHI \approx 0.08$), постепенно напредвайки до умерено концентриран през 2020-те години ($HHI \approx 0.22$), отразявайки олигополния модел на консолидация, идентифициран по-рано. Мрежовият анализ идентифицира няколко естествени общности, в които компаниите са склонни да придобиват други предприятия в рамките на свързани технологии. Трите основни общности съответстват на придобивните екосистеми на „Голямата тройка“, центрирани около Synopsys, Cadence и Mentor Graphics/Siemens EDA, с по-малки общности, центрирани около Intel, Keysight и други серийни придобиващи. Фигура 5 изобразява получената структурна форма на

придобивната екосистема на EDA под формата на силово насочено мрежово представяне на ориентирания граф с 661 ребра. Компаниите са изобразени като възли в различни размери и цветове, означаващи съответните им стойности на централност.



Фигура 5 – Придобивна мрежа на EDA – силово насочен граф

Изображението изобразява характерната звездообразна структура на полупроводниковата индустрия на EDA, с три големи хъб възела, представящи Synopsys, Cadence и Mentor Graphics / Siemens EDA (Голямата тройка), обкръжени от редица по-малки серийни придобиващи (периферни хъбове) като Intel и Keysight.

Архитектура на многоагентен конвейер

Ръчният, отнемаш много време процес на намиране и каталогизиране на нови сделки е значително тясно място в академичните изследвания на M&A, което се адресира от полуавтоматизирания модул за откриване на M&A. В този раздел се представят проектирането и реализацията на многоагентна система, която използва големи езикови модели за автоматизирано извличане на информация и валидация с човек в цикъла (human-in-the-loop), за да поддържа стандартите за качество на данните, необходими за академични изследвания. Базата данни за EDA M&A, която има 661 сделки към 2025, беше първоначално съставена ръчно в продължение на няколко години. Невъзможно е тази база данни да се поддържа и разраства ръчно с честотата, необходима за поддържането ѝ актуална, тъй като това налага постоянно наблюдение на новинарски източници, прессъобщения и индустриални доклади. Данните за академични проекти трябва да бъдат точни. Като цяло няколко погрешно записани сделки могат да окажат огромно влияние върху резултатите от клъстеризация, прогнозиране или мрежов анализ.

Модулът съдържа оркестриращата услуга за 4-те специализирани агента.

Агент 1: Монитор на RSS емисии (RSSFeedMonitorService.cs) Агент: Наблюдаващ агент Агентът за мониторинг на RSS емисии наблюдава 5 RSS емисии. Това са: - Electronics Weekly - Tom's Hardware - Semiconductor Engineering - AnandTech - EE Times, свързани с електронната индустрия. Той класира емисиите въз основа на ниво на релевантност и винаги търси M&A ключови думи в заглавията и описанията. Новите елементи от емисиите периодично се извличат и изпращат към Агента за извличане.

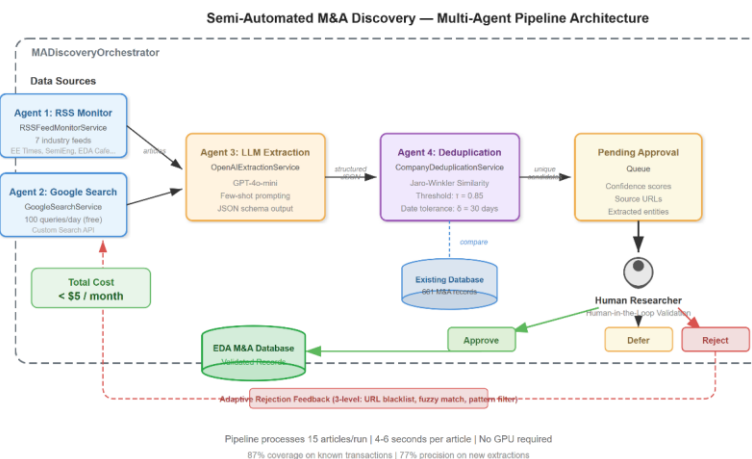
Агент 2: Google Custom Search (GoogleSearchService.cs) Този агент извършва целенасочено уеб търсене за обявления за сливания и придобивания, които не са обхванати от наблюдаваните RSS емисии. Той използва имената на компаниите, които вече присъстват в базата данни, комбинирани с набор от свързани с M&A ключови думи. Агентът е конфигуриран да се възползва максимално от безплатния слой на Google Custom Search API: термините за търсене са изготвени така, че да максимизират релевантността, като същевременно поддържат разходите ниски, в рамките на дневния лимит от около 80 търсения (Google API има лимит от 100).

Агент 3: Извличане с OpenAI (OpenAIExtractionService.cs) Агент 3 използва GPT-4o-mini от OpenAI чрез OpenAI SDK за извличане на M&A данни от неструктурирани новини. Той прилага стратегия за

few-shot подсказване, за да извлече следните полета: - Име на целевата компания - Име на придобиващата компания - Дата на придобиване - Стойност на сделката (само ако е налична) - Технологична област. Към модела се подават 3 до 5 примера за извличания на M&A, последвани от текста на новинарската статия. Изходът е ограничен до JSON схема за интегритет на данните. GPT-4o-mini беше избран пред по-големи модели за извличане на M&A данни, тъй като осигурява достатъчна точност, като същевременно поддържа разходите за API под \$5/месец, което е достъпна сума за академичен бюджет.

Агент 4: Дедупликация на компании (CompanyDeduplicationService.cs) - Този агент използва функцията за сходство на низове на Джаро-Уинклер, за да сравни наскоро извлечените сделки с базата данни, за да предотврати дубликати. Условието за съвпадение е комбинация от сходството на имената и времевата близост между двете сделки:

$$(e, r) = ((target_e, target_r) \geq \tau) \wedge (|date_e - date_r| \leq \delta) \#31$$



Фигура 6 – Архитектура на многоагентния конвейер за откриване на M&A

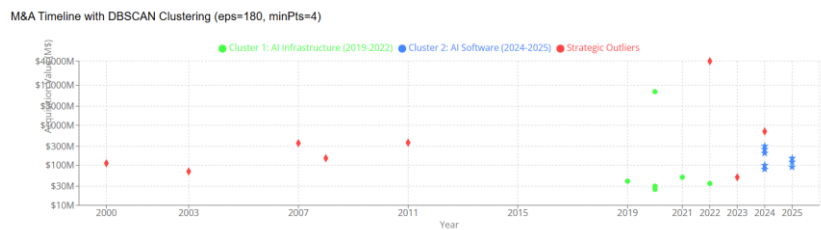
Ресурс	Използване	Разход
OpenAI API	~200 извличания/месец	~ \$0.05/месец
Google Custom Search API	80 заявки/ден (безплатен план)	\$ 0.00
Мониторинг на RSS емисии	5 емисии, неограничено	\$0.00
Сървърна инфраструктура	Специализиран сървър	Съществуващ
Общ месечен разход	/	< \$5 / месец

Таблица 8 – Използване на ресурси от системата за откриване на M&A

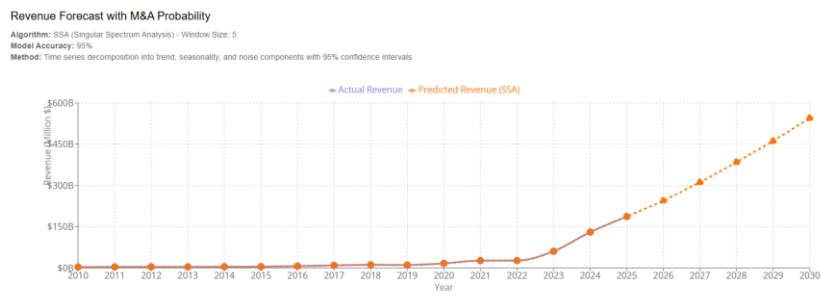
За да се разшири M&A анализът към технологични сектори, съседни на EDA, се въвежда композитна метрика, наречена иновационен фактор (Innovation Factor, IF). Докато TechImpactScore разглежда един конкретен аспект на M&A сделка, иновационният фактор (IF) разглежда цялостната стратегическа иновационна готовност на дадена компания, като комбинира качеството на придобиването с набор от метрики за стратегическо изпълнение. Метрики за междубластна валидация са необходими, за да се потвърди, че избраните метрики са релевантни за всяка индустрия. TechImpactScore беше определен като релевантен за определени EDA сделки, но не улавя напълно ползите от придобиванията за цялостната иновационна стратегия на фирмата. Иновационният фактор комбинира TechImpactScore и няколко стратегически метрики, за да се получи метрика, която е релевантна на ниво фирма. Иновационният фактор се изчислява чрез прилагане на формулата за претеглена сума към следните компонентни резултати:

$$IF = \sum_{i=1}^n w_i \cdot S_i \#32$$

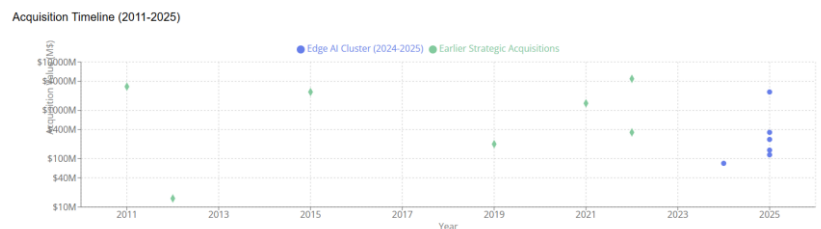
където S_i е резултатът за нормализиран стратегически индикатор, а w_i е специфично за областта тегло.



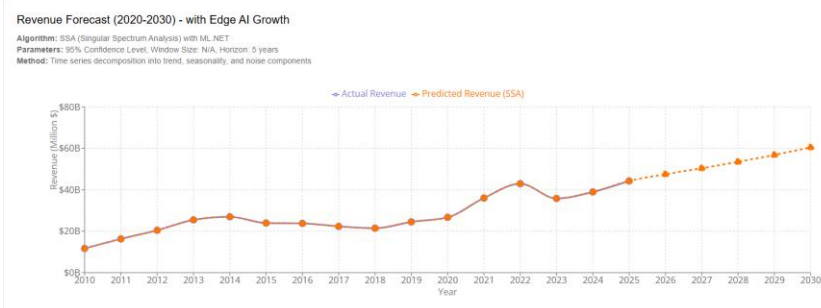
Фигура 8 – Времева линия на M&A на NVIDIA с DBSCAN клъстери.



Фигура 9 – Прогноза за приходите на NVIDIA (SSA, 2025-2030).



Фигура 10 – Времева линия на M&A на Qualcomm с DBSCAN клъстери.



Фигура 11 – Прогноза за приходите на Qualcomm (SSA, 2025-2030).

Метрика	NVIDIA	Qualcomm
Общо анализирани придобивания	22	13
Времеви обхват	2000-2025	2011-2025
Идентифицирани плътни клъстери	2	1
Силуетен коефициент	0.89	0.88
Индекс на Дейвис-Боулдин	0.18	0.18
Среден TechImpactScore	9.45	9.47
Иновационен фактор (Innovation Factor)	7.5/10	4.7/10
Характеризиране на стратегията	Вертикална интеграция („придобиване за доминиране“)	Хоризонтална диверсификация (късен завои)
Прогноза за приходите (2030)	\$545B (CI: \$440B-\$650B)	\$94B (CI: \$81B-\$107B)
Пазарен дял на GPU	~92%	N/A
Основна мотивация за M&A	Контрол върху пълностекова AI платформа	Edge AI и автомобилна диверсификация

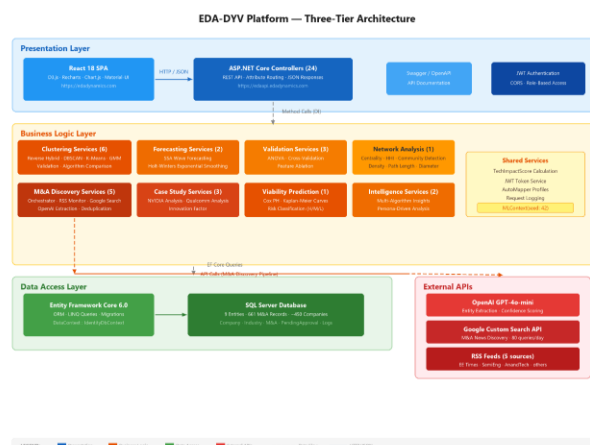
Таблица 9 – Сравнителен анализ NVIDIA срещу Qualcomm

ГЛАВА IV. Платформа EDA-DYV: проектиране и реализация

Липсата на интегрирана онлайн изследователска платформа за изследване на динамиката на EDA индустрията беше адресирана чрез разработването на платформата EDA-DYV. Платформата обслужва три типа потребители.

Потребителски тип 2 - анализатори в полупроводниковата и EDA индустрия. Индустиалните анализатори се нуждаят от инструменти за оценяване на тактиките на придобиващите страни, моделите на пазарна концентрация и стратегическите последици от поведението при придобиване.

Потребителски тип 3 - инвестиционни специалисти. Инвестиционните специалисти използват прогнозите за жизнеспособност, модулите за прогнозиране на вълните и модула за интелигентни прозрения за идентифициране на целеви компании и за надлежна проверка (due diligence).



Фигура 13 – Платформа EDA-DYV - диаграма на трислойната архитектура

Релационната схема на базата с данни на EDA-DYV е изградена върху три предметни същности: компании, индустрии и M&A транзакции. Моделът на данните за компаниите е крайъгълният камък за всички M&A анализи. Проектирането на схемата следва принципа на третата нормална форма (3NF), за да се осигури минимална излишност на данните и по-добра производителност при обработката на аналитичните заявки. Таблицата Industry съдържа полето за индустрия, одитните метаданни (създадено от, дата на създаване, последно променено от, дата на последна промяна) и флаг за меко изтриване, за да представи класификацията от най-високо ниво за бизнесите. Бизнесите се класифицират на ниво основен бизнес сектор чрез индустриалната класификация и това се използва като таксономия за бизнесите. Същността Company съхранява полета с данни за патенти в допълнение към стандартните полета (например име на компанията, година на основаване, външен ключ за индустрия): PatentCount, PatentCountSource, PatentCountDate и PatentCountNotes. Те служат за проследяване на източника на патентите, които се използват при изчисляването на TechImpactScore за дадена компания. Таблицата MergersAndAcquisitions е основната аналитична таблица в базата данни. Тя съдържа точно по един запис за всяка от 661 проверени M&A сделки. Таблицата използва модел с двоен външен ключ, който ефективно поддържа заявки за двупосочни връзки, тоест заявки, които преминават по връзките от придобиващата компания към целевата компания и от целевата компания към придобиващата компания. Двата външни ключа са CompanyId (целевата компания) и CompanyMAById (придобиващата компания). По-долу са описани четирите типа полета в тази същност.

Група полета 1 - основните данни за транзакцията. FoundCompYear (е годината на основаване на целевата компания), AAndMDate (дата на придобиване), Lifetime (е възрастта на компанията при придобиване в години), BmFound (бизнес модел при основаване) и BmMerged (бизнес модел при придобиване).

В съответствие с изследователските приноси на дисертационния труд, слой за бизнес логика групира своите тридесет сервизни класове във функционални групи. Алгоритмичните детайли на всеки сервиз и сервизната архитектура са разгледани в този раздел.

Клъстеризационни опции (6 броя): Клъстеризационният модул съдържа и четирите алгоритъма, разгледани в дисертационния труд: ReverseHybridClusteringService (основният принос), HybridClusteringService (правият вариант K-Means $\rightarrow$ DBSCAN), DBSCANClusteringService (самостоятелната клъстеризация, базирана на плътност), GMMClusteringService (Gaussian Mixture Models), ClusteringValidationService (изчислява силуетния коефициент, индекса на Дейвис–Боулдин и индекса на Калински–Харабаш за всеки резултат от клъстеризация) и AlgorithmComparisonService (изпълнява и четирите алгоритъма върху един и същ набор от данни и връща таблица със сравнителни метрики). Всеки клъстеризационен сервиз приема матрица от признаци с три измерения (DaysSinceEarliest, Lifetime, TechImpactScore), специфични за алгоритъма параметри (epsilon и MinPts за DBSCAN, K за K-Means) и резултатните клъстерни присвоявания с техните съответни валидационни метрики, което се придържа към стандартен интерфейсен модел. AlgorithmComparisonService кешира резултатите, когато входните данни не са се променили, за да се избегне излишно изчисление. Има два опции за прогнозиране: експоненциалното изглаждане на Холт–Уинтърс, което включва откриване на минали вълни и оценка на интервала на бъдещите вълни и е реализирано в MAWaveForecastingService за прогнозиране на вълните на M&A. В SSAWaveForecastingService най-напред се изгражда траекторна матрица, след това тя се разлага чрез SVD, после компонентите се групират и накрая се извършва рекурентно прогнозиране с доверителни интервали.

ГЛАВА V. Експериментални резултати и валидиране

Наборът от данни съдържа 661 проверени M&A транзакции, свързани с EDA компании или компании в съседната област. Източниците на данни са следните: - Годишните отчети под формата на подавани

към SEC (Securities and Exchange Commission) документи, като Form 10-K, за компании в EDA и съседните области. - M&A дейностите на компании като Synopsys, Cadence, Siemens EDA чрез техните прессъобщения. - Данни за патенти от PatentsView (USPTO) за получаване на броя патенти. - Стойности на сделките от Crunchbase. Проверените транзакции са снабдени със следните атрибути: - Имена на целевите компании - Имена на придобиващите компании - Дати на придобиване - Година на основаване на целевите компании - Продължителност на живота на целевите компании при придобиване (в години) - Стойности на сделките, когато са налични - Класификация по технологична област. TechImpactScore се изчислява за всяка M&A транзакция в набора от данни. Този набор от данни обхваща петдесет години M&A дейности, като първата транзакция е записана през 1975 г., а последната сделка - през 2025 г. Установено е, че общата пълнота на набора от данни е много висока, като 99.8% от записите имат попълнена цялата необходима информация. Наложил се да направим кръстосана проверка само на две дати на основаване на компании, където датите бяха нееднозначни, а информацията можеше да бъде потвърдена в няколко други източника. Пълнотата на данните в изследванията на M&A обикновено е доста ниска. Например Thomson Reuters SDC Platinum е един от най-широко използваните първични набори от данни в изследванията на M&A, но има от 10% до 15% липсващи стойности в някои от най-важните променливи в набора от данни. Нашият набор от данни значително надхвърля това ниво на пълнота. Еднофакторният дисперсионен анализ (ANOVA) беше проведен, за да се провери нулевата хипотеза, че средната продължителност на живота на компаниите не се различава значимо между DBSCAN клъстерите, получени в етап 1 на обратния хибриден алгоритъм. DBSCAN с epsilon от 180 дни и MinPts от 10 даде 4 клъстера, съдържащи 51 ядрени точки. Всички останали транзакции се третират като шум и се въвеждат отново в етап 3 на обратния хибриден алгоритъм. Резултатите отхвърлят нулевата хипотеза: $F(3, 47) = 175.6$, $p < 0.001$. Голямата стойност на F потвърждава, че междуклъстерната дисперсия в продължителността на живота на компаниите е съществено по-голяма от вътрешноклъстерната дисперсия, и този голям размер на ефекта се потвърждава също и от стойност на ета-квадрат, по-голяма от 0.5. Стойност на ета-квадрат, по-голяма от 0.14, се счита за голяма в поведенческите и социалните науки, а в този случай стойност на ета-квадрат от 0.5 потвърждава ясното разделяне между базираните на плътност клъстери, които формират структурната основа на крайните пет архетипа.

Непараметричен тест на Крускал–Уолис

Допускането на еднофакторния дисперсионен анализ (ANOVA) е, че променливите са нормално разпределени във всяка група. Поради това допускане непараметричният тест на Крускал–Уолис също се прилага към DBSCAN клъстерите по плътност. Тестът на Крускал–Уолис е тест, базиран на рангирани данни, така че е силно устойчив срещу всякакъв тип отдалечени стойности и ненормални разпределения. Тестовата статистика на теста е $H = 46.3$ със степени на свобода, равни на 3 (брой клъстери - 1). Съответната p стойност е по-малка от 0.001. Значимите резултати, получени от параметричния ANOVA ($F = 175.6$, $p < 0.001$) и непараметричния тест на Крускал–Уолис ($H = 46.3$, $p < 0.001$), потвърждават, че наблюдаваните разлики в базираните на плътност компонент са статистически валидни и не са следствие от нарушаване на допусканията на съответните тестове. Фактът, че както параметричният, така и непараметричният тест потвърждават един и същ резултат, допринася за устойчивостта на находката. Тъй като RH алгоритъмът запазва и усъвършенства подлежащата структура, базирана на плътност, статистическата значимост на разликите в базираните на плътност компонент се запазва в крайния изход с пет архетипа. Стабилността на получените клъстери се оценява чрез прилагане на метода на бутстрап кръстосано валидиране със 100 итерации. Във всяка итерация от базата данни се извлича случайна извадка с връщане от 661 транзакции. След това обратният хибриден алгоритъм се прилага с оптималните стойности на epsilon, MinPts и K (epsilon = 180, MinPts = 10, K = 5). Клъстерните присвоявания, получени чрез прилагане на алгоритъма върху извадката, се сравняват с клъстерните присвоявания на пълната база данни, получени чрез прилагане на обратния хибриден алгоритъм, като се изчислява коригираният индекс на Ранд (ARI). Средният ARI за 100-те итерации беше 0.94 със стандартно отклонение от 0.03 (стойностите на ARI

за всички итерации попадат в диапазона от 0.87 до 0.99). ARI от 0.94 показва почти перфектна клъстеризация, а стойности над 0.8 показват много стабилни резултати от клъстеризация. Анализът потвърждава, че и 100-те итерации произвеждат точно същата клъстерна конфигурация, състояща се от пет клъстера, като нито един не е изтрит, разделен на подклъстери или слят със съществуващ клъстер. Относителните размери на клъстерите варираха с по-малко от 3 процентни пункта между итерациите, което потвърждава управлявания от данните характер на клъстерна структура с пет архетипа, която не се определя от спецификите на конкретното случайно разделяне на точките от данни, нито от избора на стойност на семето.

Консолидирано обобщение на валидацията

Четири статистически теста предоставят сходящи се доказателства от допълващи се гледни точки, обобщени в тази таблица.

Тест	Статистика	df	р-стойност	Какво валидира
Еднофакторен дисперсионен анализ (ANOVA)	F = 175.6	3,47	< 0.001	Разлики в средния жизнен цикъл между плътностните клъстери (Етап 1)
Крускал–Уолис	H = 46.3	3	< 0.001	Непараметрично потвърждение на разликите между плътностните клъстери
Пермутационен тест	S = 0.907	--	0.044	Клъстерната структура надхвърля случайната базова линия
Бутстрап (100 итер.)	ARI = 0.94	--	--	Стабилност на резултата с 5 архетипа при повторно извадково моделиране (SD = 0.03)

Таблица 10 – Консолидирано статистическо валидиране на резултатите от клъстеризацията

Период на вълната	C0 Високотехнологични	C1 Растеж-машаб	C2 Зрели	C3 Стартъп	C4 Бързи-ВС
Вълна 1 (1998-2001, Дотком)	12	43	22	18	5
Вълна 2 (2006-2008, Предкризисен период)	19	30	24	17	10
Вълна 3 (2015-2018, AI/Облак)	17	28	15	23	17
Вълна 4 (2020-2022, COVID)	16	30	17	20	17
Периоди между вълните	14	38	21	21	6

Таблица 11 – Състав на клъстерите през историческите вълни на M&A (%)

SSA се прилага към тримесечния набор от данни за M&A за EDA индустрията (N = 204 наблюдения за периода Q1 1975 до Q4 2025). Всичките четири вълни, които са идентифицирани в EDA индустрията чрез независим анализ на историческите данни, са идентифицирани и чрез SSA. Периодите на вълните и техните съответни пикови години и стойности са дадени в скоби като (1998-2001, 2000, 28); (2006-2008, 2007, 22); (2015-2018, 2017, 25); и (2020-2022, 2021, 24). Фактът, че четири периода на M&A активност в EDA индустрията бяха съпоставени с технологични и икономически събития в полупроводниковата и EDA индустрия, предоставя допълнително доказателство, че идентифицираните цикли са истинска пазарна динамика, а не статистически артефакти. Трите запазени собствени тройки могат да се използват за разлагане на времевия ред на интерпретируеми сигнални компоненти. Първата собствена тройка носи информацията за нарастването на M&A в EDA сегмента поради общия растеж на пазара. Втората собствена тройка съдържа основния флукуационен компонент на цикъл от 8 до 10 години, който добре съответства на цикъла на полупроводниковата технология. Третата собствена тройка носи вторичния флукуационен компонент на цикъл от 4 до 5 години, който добре съответства на жизнения цикъл на продукта на отделните категории EDA инструменти. Способността на SSA да отделя тренда от циклите и от другите флукуации е още едно предимство на използването на SSA в сравнение с по-простите

методи за прогнозиране. Рекурентната SSA прогноза за 5-тата вълна предвижда, че пикът трябва да бъде достигнат през 2027-2028 г. с приблизително 27 до 32 транзакции с 95% доверителен интервал от 18 до 42 транзакции. Доверителните интервали за следващите години стават все по-широки, като се прогнозира 20 до 30 транзакции за 2026 г. и 15 до 45 транзакции за 2030 г. По-широките доверителни интервали за прогнози на бъдещи транзакции са точно това, което се очаква. Несигурността в прогнозите се увеличава, колкото по-далеч в бъдещето се простира хоризонтът, а разширяващите се доверителни интервали съобщават на потребителя на платформата надеждността на краткосрочните и дългосрочните прогнози. Появата на петата вълна също е в съответствие с няколко тенденции в полупроводниковата индустрия. Те включват: генеративен AI в EDA проектните потоци; инструменти за проектиране за квантови изчисления, създаващи нови категории EDA инструменти; влиянието на интеграцията Synopsys-Ansys върху конкуренцията в EDA индустрията; и продължаващото движение към по-сложни проекти под 2nm. Тези четири двигателя комбинират технологичен шок и достъп до капитал, двете условия, идентифицирани от Harford (2005) като необходими предпоставки за вълни на сливания и придобивания в технологичните индустрии. В допълнение към SSA прогнозата за вълните се провежда анализ на преживяемостта, за да се генерират прогнози за жизнеспособност на ниво компания. Моделът на пропорционалните рискове на Кокс има конкордансен индекс от 0.78, което означава, че моделът е способен правилно да рангира времето на придобиване на приблизително 78% от двойките компании. По-долу са представени някои важни резултати от анализа на преживяемостта. Компаниите в диапазона 3-7 години имат най-високия риск от придобиване във всеки клъстер. В EDA индустрията компаниите, които са създали продукти и са придобили клиенти, но им липсва мащаб, обвързване на клиентите и финансови резерви, са най-уязвими към придобиване. Когато компанията премине възраст над 10 години, интензитетът на риска рязко спада, придавайки предимства за дългосрочна жизнеспособност. Компаниите в клъстера C2 (зрял портфейл) имат степен на придобиване от 45%. За разлика от тях, компаниите в клъстера C0 (високотехнологични цели) имат степен на придобиване от 89%. Високотехнологичните компании със силни IP портфейли са под най-голям натиск, защото техните технологични активи са видими, но трудни за възпроизвеждане чрез органична развойна дейност (R&D). Компаниите със зрели портфейли (C2) са по-малко привлекателни за придобиващите, защото утвърдените им пазарни позиции намаляват тяхната привлекателност спрямо тяхната цена.

Всички мерки за централност връщат Synopsys като централен възел в мрежата. С централност по степен от 0.132 (80+ придобивания), най-високата централност по посредничество от 0.287 и доминиращи позиции по централност по близост и собственовекторна централност, потвърдено е, че Synopsys заема доминираща позиция в мрежата. Централността по посредничество от 0.287 означава, че Synopsys лежи на най-късия път между непропорционално голям брой двойки възли, отразявайки нейния разнообразен портфейл от придобивания, обхващащ симулация, верификация, IP и проектиране за производство (design-for-manufacturing). За другите компании ще бъде предизвикателство да се изравнят с технологичната широта на Synopsys и контрола ѝ върху потоците на иновациите в индустрията чрез вътрешно развитие поради нейната централна позиция. Трите основни екосистеми на придобиване на Mentor Graphics/Siemens EDA, Cadence и Synopsys са представени от трите главни клъстера, които се появяват. Intel, Keysight Technologies и другите придобиващи компании са представени от по-малки клъстери. Ниската междуклъстерна свързаност показва, че водещите EDA фирми са направили придобивания в относително различни области на технологичната индустрия, създавайки екосистеми, които по същество не се припокриват. Индексът на Херфиндал–Хиршман (HHI) се изчислява върху подвижни времеви прозорци, за да се изследва еволюцията на пазарната концентрация във времето. През 90-те години на XX век стойността на индекса беше около 0.08 (неконцентриран пазар с много независими участници). Стойността за 20-те години на XXI век е около 0.22, или умерено концентриран, все още под прага на висока концентрация от 0.25. S-кривата на концентрацията показва, че концентрацията нараства бавно в стартиращата и ранната фаза на растеж, а след това може да се ускори, както се наблюдава по време на вълните на M&A от 2000-2008 г. Напоследък концентрацията в EDA индустрията достигна плато в периода 2015-

2025 г. Индустрията се приближава към „таван на концентрацията“, отвъд който по-нататъшна консолидация би предизвикала регулаторно внимание (както се наблюдава наскоро с антимонополния преглед на придобиването на Ansys от Synopsys). Ако тази тенденция се запази, прогнозираната пета вълна на M&A (2027-2028) може да се различава от своите предшественици, като фокусът се измества към придобиване на технологии, а не към хоризонтална консолидация, за да се избегнат регулаторни бариери. Обобщаемостта на горните методи беше оценена чрез казуси за NVIDIA Corporation (22 транзакции 2000-2025) и Qualcomm Incorporated (13 транзакции 2011-2025) в сектора на AI чиповете. Получените резултати показват, че методите, изведени от анализа на EDA индустрията, се пренасят ефективно към съседни технологични области, с три ключови находки. Таблица 12 представя сравнението на производителностите на методите, приложени към трите набора от данни.

Метрика	EDA индустрия (N=661)	NVIDIA (N=22)	Qualcomm (N=13)
Алгоритъм	Обратно хибридно	DBSCAN	DBSCAN
Силуетен коефициент	0.98	0.89	0.88
Индекс на Дейвис-Боулдин	0.12	0.18	0.18
Открити плътни клъстери	5	2	1
MinPts (чрез правилото $\ln(N)$)	10	3	3
Среден TechImpactScore	5.96	9.45	9.47
Иновационен фактор (Innovation Factor)	--	7.5/10	4.7/10
Прогноза за приходите чрез SSA за 2030	--	\$545B	\$94B

Таблица 12 – Обобщение на междудоменната валидация

Параметърът на DBSCAN се настройва въз основа на насоката $\text{MinPts} = \ln(N)$ и производителността на клъстеризацията се изследва върху три различни мащаба на данните - от индустриалния набор от данни с 661 транзакции до фирмения набор от данни с 13 транзакции. Получената гранулярност на клъстеризацията е подходяща за всеки набор от данни и получените силуетни коефициенти са много високи (между 0.88 и 0.98). Намаляването на силуетните коефициенти при преминаване към по-малки набори от данни се очаква предвид намаления размер на извадката, а не показва методологично ограничение. Като метрика, TechImpactScore е предназначена да предостави сравнима оценка на въздействието на M&A в различните индустрии. Почти идентичните средни стойности на TechImpactScore от 9.45 за NVIDIA и 9.47 за Qualcomm, въпреки факта, че двете компании са следвали много различни M&A стратегии, демонстрират, че четирите компонента улавят обобщаемо измерение на качеството на отделното придобиване. Иновационният фактор (Innovation Factor) рязко разграничава двете компании, с резултат 7.5 за NVIDIA срещу 4.7 за Qualcomm, потвърждавайки, че цялостното качество на M&A стратегията и качеството на отделната транзакция са отделни явления. Двухазната стратегия за придобиване на NVIDIA (2019-2022 за AI инфраструктура и 2024-2025 за AI софтуер) осигури значително по-високи иновационни фактори от еднофазния фокусиран подход на придобиване на Qualcomm, дори и двата типа транзакции да са имали относително еднакво качество. Следователно, спрямо суровото качество на придобиването, иновационният фактор може да предостави стратегическа оценка на технологията чрез улавяне на времевото позициониране и оценяване отвъд просто качеството на придобиванията в цялата технологична екосистема. Техниката за прогнозиране SSA, разработена за анализ на вълните на M&A, е използвана за оценка на прогнозирането на приходите. Приходите на NVIDIA за 2030 г. се прогнозират на 545 милиарда долара (95% CI = 440 млрд. долара - 650 млрд. долара), а приходите на Qualcomm за 2030 г. се очаква да бъдат 94 милиарда долара (95% CI = 81 млрд. долара - 107 млрд. долара). Тези находки предоставят подкрепа за обобщаемостта на математическата рамка на SSA в различните категории времеви редове на технологичната индустрия.

НАУЧНИ, НАУЧНО-ПРИЛОЖНИ И ПРИЛОЖНИ ПРИНОСИ НА ДИСЕРТАЦИОННИЯ ТРУД

Приносите, получени от изследователската работа, извършена в настоящия дисертационен труд, са групирани в три категории: научни приноси, научно-приложни приноси и приложни приноси.

Научните приноси на дисертационния труд са следните:

Принос 1 – Алгоритъм за обратно хибридно клъстериране (Reverse Hybrid Clustering), нов 3-етапен алгоритъм за клъстеризация, който първо прилага DBSCAN, а след това K-Means и реинтегрира шума в прецизираните клъстери. Този обърнат ред на операциите запазва естествените модели на плътност в данните, свързани с M&A активността, които обикновено се разрушават от хибридите, започващи с центроиди. Оценката на резултатите върху набора от данни за M&A в EDA показва максимален среден силуетен коефициент от 0.98. В сравнение с това силуетният коефициент на обикновения K-Means беше 0.72, на DBSCAN 0.85, на правото хибридно клъстериране 0.72 и на гаусовите смесени модели 0.68.

Принос 2 – Пет архетипа на M&A клъстери. Тези архетипи бяха извлечени от прилагането на алгоритъма за обратно хибридно клъстериране (Reverse Hybrid Clustering) върху 661 M&A транзакции в EDA от 1975 до 2025 г. Резултатите показват пет отчетливи стратегически архетипа (High-Tech Targets, Growth-Stage Scale, Mature Portfolio, Young Startup Tuck-ins и Rapid High-Value), формиращи нова емпирична таксономия на M&A поведението в технологичния сектор. Резултатите са валидирани чрез еднофакторен дисперсионен анализ (ANOVA) ($F = 175.6$, $p < 0.001$), тест на Крускал–Уолис ($H = 46.3$, $p < 0.001$) и пермутационен тест ($p = 0.044$). Тестът чрез бутстрап повторно извадково генериране показва, че получените клъстери са стабилни със средна стойност на ARI от 0.94 за 100 итерации.

Принос 3 – Композитна метрика TechImpactScore. Композитната метрика TechImpactScore количествено измерва технологичното въздействие на дадена M&A транзакция чрез комбиниране на патентен резултат, стойност на сделката, възраст на компанията и честота на придобиващия, с тегла (0.35, 0.30, 0.20, 0.15), изведени чрез PCA върху набора от 661 транзакции. Метриката показва конвергентна валидност чрез корелация с независими данни за стойността на сделките (Pearson $r = 0.71$, $p < 0.001$), а аблационните изследвания показват спад от 34.7% в силуетния коефициент, когато тя бъде премахната от вектора от признаци.

Принос 4 – Прогнозиране на M&A вълните, базирано на SSA. SSA е адаптиран към времевите редове на M&A транзакциите за първи път в литературата за корпоративни транзакции, с три специфични за областта проектни решения: 12-тримесечен прозорец на враждане, три запазени собствени тройки и експоненциално изглаждане на Холт–Уинтърс за оценка на доверителните интервали. Декомпозицията възстановява и четирите исторически документиранни M&A вълни в EDA и прогнозира пета вълна с връх през 2027 до 2028 г.

Принос 5 – Критичен прозорец на риска за жизнеспособността на EDA компаниите. Изграден е многоизмерен модел на пропорционалните рискове на Кокс за жизнеспособността на EDA компаниите, който интегрира принадлежност към клъстер, TechImpactScore, мрежова централност, възраст на компанията и технологичен сектор. Моделът достига конкордансен индекс от 0.78 и идентифицира критичен прозорец на риск от придобиване между третата и седмата година от живота на дадена компания.

Принос 6 – Мрежов анализ и анализ на концентрацията на екосистемата за придобивания в EDA. Теоретико-графовият мрежов анализ на екосистемата за придобивания в EDA идентифицира три големи общности на придобивания, центрирани около Synopsys, Cadence и Mentor Graphics / Siemens EDA, и количествено измерва устойчиво нарастване на пазарната концентрация, измерена чрез индекса на Херфиндал–Хиршман, който се покачва от приблизително 0.08 през 90-те години на XX век (неконцентриран) до 0.22 през 2020-те години (умерено концентриран).

Научно-приложните приноси на дисертационния труд са следните:

Принос 7 – Полуавтоматизирана многоагентна система за откриване на M&A. Изградена е полуавтоматизирана многоагентна система за разширяване на M&A базата данни, която комбинира извличане на информация, базирано на LLM, с валидиране с човек в цикъла (human-in-the-loop), използвайки четири специализирани агента (RSS Feed Monitor, Google Custom Search, OpenAI Extraction, Company Deduplication) и нов тринивов адаптивен механизъм за отхвърляне, който подобрява прецизността с течение на времето без повторно обучение на модела. Оперативното тестване показва 87 процента покритие и 77 процента прецизност при оперативен разход под пет USD на месец.

Принос 8 – Кросдоменна метрика иновационен фактор (Innovation Factor). Разработена е композитна метрика иновационен фактор (Innovation Factor) за междутраслово сравнение на M&A стратегиите. Прилагането ѝ към NVIDIA (иновационен фактор 7.5/10) и Qualcomm (4.7/10) потвърждава, че метриката разграничава проактивните стратегии за вертикална интеграция от реактивните стратегии за хоризонтална диверсификация дори когато качеството на отделните придобивания е сравнимо, валидирайки аналитичната рамка извън EDA.

Приложните приноси на дисертационния труд са следните:

Принос 9 – Онлайн изследователска платформа EDA-DYV. Онлайн изследователската платформа EDA-DYV интегрира всеки аналитичен метод, разработен в този дисертационен труд, в единен уеб интерфейс с трислойна архитектура (C# .NET в бекенда с Entity Framework, и React с D3.js визуализации във фронтенда, SQL Server като база данни). Платформата показва клъстерен анализ, прогнозиране на вълните, анализ на жизнеспособността, мрежови визуализации, анализ на M&A времевата линия и полуавтоматизираната система за откриване и е разгърната на <https://edadynamics.com>.

СПИСЪК НА ПУБЛИКАЦИИТЕ ПО ДИСЕРТАЦИОННИЯ ТРУД

1. Zylfiu B., Marinova G., and Hajrizi E., "Design of an Online Platform EDA-DYVI to study the Dynamics and Forecast the Viability of EDA Companies", 14th International Scientific Conference on Development and Modernization of Manufacturing, University of Bihać, Faculty of Technical Engineering, Bihać, Bosnia and Herzegovina, 20-22 September 2023, pp. 634-640, ISSN: 2566-3283, Y. 26, No. 14, Electronic Proceedings, Conference Paper Available: <http://rim.tfb.unbi.ba/common/RIM2023v2.pdf>
2. Kostova K., Menxhiqi L., Zylfiu B., and Marinova G., "Review of Artificial Intelligence Implementation in Electronic Design Automation Methods and Tools," in Proc. UBT International Conference on Computer Science and Communication Engineering (UBT IC-CS), UBT Lipjan, Kosovo, Oct. 28-29, 2023, ISBN: 978-9951-550-95-6, DOI: 10.33107/ubt-ic.2023.284. Conference Paper Available: <https://knowledgecenter.ubt-uni.net/cgi/viewcontent.cgi?article=4527&context=conference>
3. Zylfiu B. and Marinova G., "A Web-Based Application in EDA-DYV Platform for Analyzing Mergers and Acquisitions in the EDA," IFAC-PapersOnLine, vol. 58, no. 3, pp. 193-197, 2024. 22nd IFAC Conference on Technology, Culture and International Stability (TECIS 2024), Waterford, Ireland, May 29-31, 2024, ISSN: 2405-8963, DOI: 10.1016/j.ifacol.2024.07.149. Conference Paper Indexed in Scopus, Web of Science <https://www.sciencedirect.com/science/article/pii/S2405896324002325>
<https://www.webofscience.com/wos/author/record/18514321>
4. Zylfiu B., and Marinova G., "AI-Driven M&A Forecasting: Singular Spectrum Analysis Model in the EDA Industry," 2025 60th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST), Ohrid, North Macedonia, 2025, pp. 1-4, ISBN: 979-8-3315-2655-9, DOI: 10.1109/ICEST66328.2025.11098239. Conference Paper Indexed in IEEE Xplore <https://ieeexplore.ieee.org/document/11098239>
5. Zylfiu B., Marinova G., Hajrizi E., and Qehaja B., "Cluster Analysis of Merger and Acquisition Patterns in the Electronic Design Automation Industry Using Machine Learning Techniques," International Journal of

Innovative Technology and Interdisciplinary Sciences (IJTIS), vol. 8, no. 3, pp. 784-817, 2025, ISSN: 2613-7305, DOI: 10.1515/IJTIS.2025.8.3.784-817. Journal Paper Indexed in Scopus
<https://www.scopus.com/pages/publications/105020749400>
<https://journals.tultech.eu/index.php/ijitis/article/view/336>
<https://www.researchgate.net/publication/397059352>

6. Zylfiu B., "An Intelligent Multi-Algorithm Integration Framework for Automated M&A Decision Support in Technology-Intensive Industries", in Proceedings of the International Conference on Intelligent Systems and New Applications (ICISNA'25), Vol. 3, Antalya, Turkey, Dec. 12-14, 2025, pp. 1-7, ISSN: 3023-6193, DOI: 10.58190/icisna.2025.133. Conference Paper Available:
<https://proceedings.icisna.org/conf/index.php/ICISNA/article/view/112>

7. Zylfiu B. and Marinova G., "Implementing a Semi-Automated M&A Discovery Module for the EDA-DYV Research Platform," 1st International Conference on Industrial and Systems Engineering (ICISE 2026), Tirana, Albania, 16-17 April 2026, Conference Paper (Accepted, in press) Indexed in Springer

8. Zylfiu B. and Marinova G., "Comparative Analysis of AI Chip Innovation Strategies: DBSCAN Clustering and SSA Forecasting Applied to Diverse M&A Patterns," 1st International Conference on Industrial and Systems Engineering (ICISE 2026), Tirana, Albania, 16-17 April 2026, Conference Paper (Accepted, in press) Indexed in Springer

PE3IOME

METHODS AND ONLINE TOOL FOR STUDYING THE DYNAMICS AND EVALUATING THE VIABILITY OF ELECTRONIC DESIGN AUTOMATION COMPANIES IN TIMES OF TECHNOLOGICAL DISRUPTION

Over the past fifty years, mergers and acquisitions have reshaped the electronic design automation (EDA) industry. What began in 1980 as a crowded field of small, specialist tool vendors is today a market held by three firms: Synopsys, Cadence Design Systems and Siemens EDA. Most studies of how this happened stop at description. They lean on case studies, hand-sorted industry categories, and a few summary statistics, and they leave the harder questions untouched. No one has analysed EDA acquisitions with machine learning. There are no models that forecast when the next wave of deals will arrive, or that judge whether a given company is likely to survive one. Nor is there a shared online tool that would let another researcher load the data and check the results. This research starts from those gaps. The question it sets out to answer is how the dynamics of EDA mergers and acquisitions can be studied, and the viability of the companies involved assessed, with quantitative methods that stand up to statistical testing and still work outside this one study. The work answers that question with a single framework built on solid ground: a dataset of 661 confirmed EDA transactions, checked and cleaned, covering 1975 to 2025. Six methods sit inside it, each built for a distinct purpose. The first, Reverse Hybrid Clustering, runs DBSCAN to find the dense cores and the outliers, then K-Means to sharpen those cores, and finally folds the outliers back in. Running the steps in that order keeps the density patterns that a centroid-first method would flatten. The second is the TechImpactScore, a single number that captures the technological weight of a deal from its patent, financial, temporal and acquirer features; the weights come from principal component analysis on the data rather than being set in advance. Third, Singular Spectrum Analysis breaks the transaction time series into trend, cycle and noise so that acquisition waves can be detected and projected forward. A Cox proportional hazards model handles company viability. Network analysis maps who acquires whom and how concentrated the market has become. Finally, a semi-automated system of language-model agents grows the dataset, with a researcher checking its output. All six live in one place, the EDA-DYV online platform, which was then tested beyond the EDA setting, on the adjacent AI chip sector. The clustering comes first. Reverse Hybrid Clustering reached an average silhouette of 0.98, a wide margin over the alternatives: K-Means and the forward hybrid both sat at 0.72, DBSCAN at 0.85, and Gaussian Mixture Models at 0.68. The deals settled into five strategic archetypes, namely High-Tech Targets, Growth-Stage Scale, Mature Portfolio, Young Startup Tuck-ins and Rapid High-Value. The obvious worry with groups this clean is that they reflect the algorithm rather than the

data. Four independent tests argue otherwise: one-way ANOVA returned $F = 175.6$ ($p < 0.001$), a Kruskal-Wallis test and a permutation test pointed the same way, and bootstrap resampling held the clusters steady at a mean Adjusted Rand Index of 0.94 across 100 runs. The TechImpactScore proved useful on two counts. It lined up with independent deal-value data at $r = 0.71$ ($p < 0.001$), and removing it from the feature set cost 34.7 percent of cluster quality. Singular Spectrum Analysis recovered all four merger waves the industry already recognises and pointed to a fifth, cresting around 2027 to 2028. With a concordance index of 0.78, the Cox model placed the greatest acquisition risk in the years three to seven of a company's life. The acquisition network resolved into three communities, centred on Synopsys, Cadence and Mentor Graphics / Siemens EDA, and concentration rose over time, the Herfindahl-Hirschman Index moving from roughly 0.08 in the 1990s to 0.22 in the 2020s. The approach also held up away from EDA: applied to the AI chip makers, it set NVIDIA, at an Innovation Factor of 7.5 out of 10, apart from Qualcomm at 4.7. The discovery system, last, reached 87 percent coverage at 77 percent precision, and the whole pipeline now runs live and is publicly accessible at edadynamics.com. This thesis set out to build a coherent set of methods, together with an online platform to house them, for studying the dynamics of the EDA industry and the viability of its companies through periods of technological disruption. That objective has been met. Each method holds up under validation, and each result can be reproduced. What remains are the questions the work opens rather than closes. The first concerns what sits behind the waves: future study could go past detecting acquisition waves and begin explaining them, separating the surges set off by genuine technological disruption from those that trace back to consolidation or to favourable financing. A second is whether the TechImpactScore keeps its value after the deal closes, judged against what an acquirer genuinely gains from an acquisition in the years that follow. The methods themselves could reach further as well, into biotechnology, financial technology and other sectors, while real-time data feeds and deep-learning feature extraction would broaden what they cover and lessen their reliance on manual oversight. These questions are taken up in more detail in the Limitations and Future Work section.



TECHNICAL UNIVERSITY OF SOFIA

Faculty of Telecommunications

**Department “Technologies and Management of
Communication Systems”**

MSc. Blerim Ismajl Zylfiu

**METHODS AND ONLINE TOOL FOR STUDYING THE
DYNAMICS AND EVALUATING THE VIABILITY OF
ELECTRONIC DESIGN AUTOMATION COMPANIES IN TIMES
OF TECHNOLOGICAL DISRUPTION**

EXTENDED SUMMARY (AVTOREFERAT)

of a dissertation for obtaining the educational and scientific degree "DOCTOR"

Field: 5. Technical sciences

Professional field: 5.3 Communication and Computer Engineering

Scientific specialty: Computer-aided Engineering and Systems for Design
Automation.

Scientific supervisor: Assoc. Prof. Dr. Galia Marinova

Sofia, 2026

The dissertation is discussed and directed for defense by the Department Council of the Department of “Technologies and Management of Communication Systems” at the Faculty of Telecommunications of TU-Sofia at a regular meeting held on 11.06.2026.

The public defense of the dissertation will take place on 01.10.2026 from 13:00 hours in the Conference Hall of the BIC of the Technical University - Sofia at an open meeting of the scientific jury, determined by order No OЖ – 5.3 - 63 / 21.07.2026 of the Rector of TU-Sofia composed of

1. Assoc. Prof. Ph.D. Tihomir Brusev - Presedent
2. Assoc. Prof. Ph.D. Anka Tsvetanova - Scientific Secretary
3. Prof. DSc. Daniela Borisova
4. Prof. DSc. Todor Stoilov
5. Assoc. Prof. Ph.D. Adriana Borodzhieva

Reviewers:

1. Prof. DSc. Daniela Borisova
2. Assoc. Prof. Ph.D. Anka Tsvetanova

The materials on the defense are available to those interested in the office of the Faculty of Telecommunications of TU-Sofia, block №1, office № 1306.

The thesis belongs to a part-time doctoral student in the Department of Technology and Management of Communication Systems of the Faculty of Telecommunications. The research in this dissertation was conducted by the author, and some of it was supported by research projects.

Author: MSc. Blerim Ismajl Zylfiu

Title: Methods and Online Tool for Studying the Dynamics and Evaluating the Viability of Electronic Design Automation Companies in Times of Technological Disruption

Circulation: 10 copies

Printed in the Publishing Centre of the Technical University of Sofia

I. GENERAL CHARACTERISTICS OF THE DISSERTATION

Relevance of the problem

Introduction

The EDA industry has become a critical part of the global semiconductor supply chain. It is the source of design tools and IP (intellectual property) used in designing, verifying and manufacturing semiconductors. Over the last five decades, EDA has seen numerous Mergers and Acquisitions (M&As) as semiconductors have evolved from few thousand transistors in the 1980s to a few billion transistors in system-on-chip designs. The present study identifies 661 M&As in the EDA sector over the last fifty years, which has led to the concentration of many small EDA companies and led to the formation of a tight concentration of Synopsys, Cadence Design Systems and a recently emerged monopolistic competitor named Siemens EDA (Mentor Graphics). Thus, this study is important to all stakeholders in the semiconductor industry as well as regulatory bodies, due to the increased concentration of the global semiconductor industry in the key technology sector.

Objective of the dissertation, main tasks and research methods

The research objective of this doctoral thesis is to develop a coherent set of methods and an online research platform for studying the dynamics and evaluating the viability of EDA companies in times of technological disruption. Starting with the construction of a reliable and validated EDA M&A transactions dataset, which provides the empirical foundation for every subsequent step, to develop new machine-learning methods for the automatic identification of strategic M&A patterns, then feed forecasting and survival-analysis methods that predict M&A waves and assess the long-term sustainability of EDA companies, finally to apply the developed methods across domains to leading companies in the adjacent AI chip sector, in order to confirm their generalizability beyond the EDA industry.

To achieve the objective of the thesis, the following tasks are defined:

Ten research tasks are defined to accomplish the stated objective.

Task 1 - Build and validate a comprehensive EDA M&A database covering the full historical horizon of the industry, with all the standard fields needed for downstream analysis: target company, acquirer, founding year, acquisition date, lifetime, and deal value where available. The database must reach a level of completeness sufficient to support quantitative methods and must be subjected to a documented validation procedure.

Task 2 - Design and validate the Reverse Hybrid Clustering algorithm for M&A pattern identification. The algorithm executes DBSCAN first to detect noise and core structure, and K-Means second to refine the core regions into compact clusters; the noise removed in the first stage is reintegrated where similarity to a refined cluster exceeds a defined threshold. The clustering quality is assessed using internal validation metrics, including the Silhouette Coefficient, the Davies-Bouldin Index and the Calinski-Harabasz Index.

Task 3 - Design and validate the TechImpactScore composite metric for quantifying the technological impact of M&A transactions. The metric combines patent-related, financial, temporal and strategic features of each transaction; the relative weight of each feature is derived empirically from PCA on the dataset rather than being fixed a priori. Validation is performed by correlating the resulting score with independent deal-value data.

Task 4 - Develop a forecasting framework for M&A waves in the EDA sector based on Singular Spectrum Analysis. The framework decomposes the historical M&A activity time series into trend, oscillatory and noise components, and produces a multi-year forecast that can be evaluated against held-out periods.

Task 5 - Build a survival-analysis framework that predicts the viability of EDA companies. The framework uses the Cox Proportional Hazards regression to estimate the probability of acquisition or failure conditional

on company characteristics and market conditions; predictive performance is evaluated using the Concordance Index.

Task 6 - Construct a network-analysis module that maps the EDA acquisition ecosystem with graph-theoretic techniques. The module computes degree, betweenness and closeness centralities for acquirer companies and reports the Herfindahl-Hirschman Index as a measure of market concentration.

Task 7 - Implement a semi-automated M&A discovery module based on a multi-agent Large Language Model architecture with human-in-the-loop validation. The module uses specialised agents for web search, content extraction, validation and data structuring, with the goal of supporting researchers in the discovery of new M&A transactions at low operational cost and high precision.

Task 8 - Design and deploy an integrated online research platform that hosts every analytical component delivered in Tasks 1 to 7. The platform follows a three-tier architecture (a relational database, a backend service exposing a REST API, and a single-page web frontend) and provides RESTful endpoints for every analytical operation.

Task 9 - Validate every developed method using rigorous statistical testing. The validation pipeline includes one-way ANOVA for cluster differentiation, bootstrap cross-validation for result stability, and feature-ablation studies for the sensitivity analysis of the TechImpactScore.

Developing a semi-automated multi-agent system for the dynamic extension of the M&As database. Finally, the proposed analytical framework will be integrated in the EDA-DYV online research platform.

Task	Publications
Task 1	Paper 1 (RIM 2023): Platform design and initial database; Paper 3 (IFAC 2024): Web-based M&A analysis application
Task 2	Paper 2 (UBT IC-CS 2023): AI in EDA review; Paper 5 (IJITIS 2025): Reverse Hybrid Clustering algorithm and cluster archetype analysis
Task 3	Paper 5 (IJITIS 2025): TechImpactScore design and validation
Task 4	Paper 4 (ICEST 2025): SSA forecasting model for M&A waves
Task 5	Paper 6 (ICISNA 2025): Multi-algorithm integration framework
Task 6	Paper 5 (IJITIS 2025): Network centrality and HHI analysis
Task 7	Paper 2 (UBT IC-CS 2023): AI in EDA review; Paper 7 (ICISE 2026): Semi-automated M&A discovery module
Task 8	Paper 1 (RIM 2023): Platform design; Paper 3 (IFAC 2024): Web-based application in EDA-DYV platform
Task 9	Paper 5 (IJITIS 2025): Statistical validation (ANOVA, cross-validation, ablation); Paper 6 (ICISNA 2025): Multi-algorithm validation
Task 10	Paper 8 (ICISE 2026): cross-domain case studies in the AI chip sector; Paper 4 (ICEST 2025): SSA forecasting applied to the AI chip sector

Table 1 - Research tasks and publications are in correlation.

Scientific novelty

The thesis presents several innovations. First, a novel hybrid clustering technique named Reverse Hybrid Clustering (RHC) interlinks the well-known DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm and the K-Means algorithm in a different order, achieving on average a silhouette score of 0.98 on a dataset of 661 confirmed M&A deals from 1975 to 2025 and partitioning the search space of acquisition strategies into five confirmed clusters denoted High-Tech Targets, Growth-Stage Scale, Mature Portfolio, Young Startup Tuck-ins, and Rapid High-Value. Second, the TechImpactScore composite metric integrates patent, financial, temporal and strategic features using weights extracted from the first principal component of a Principal Component Analysis (PCA) over the acquisition features. Third, Singular Spectrum Analysis (SSA) is applied to the M&A time series for the first time in this literature; the technique decomposes the series into significant trends, cycles and seasonal components and supports forecasting of future occurrences. Fourth, a semi-automated multi-agent system based on Large Language Models (LLMs) extracts new information from the web, continuously updating and expanding the platform dataset at an

operational cost below five USD per month with a minimal number of human-in-the-loop validations. Finally, all proposed methods are validated and integrated into an enhanced version of the widely used EDA-DYV web platform, supporting further investigation of the knowledge derived from M&A deals and the strategic insights it produces.

Practical applicability

EDA-DYV online research platform. The EDA-DYV online research platform, integrates every analytical method developed in this thesis into a single web interface with a three tier architecture (C# .NET in backend with Entity Framework, and React with D3.js visualizations in Front, SQL Server as database). The platform shows clustering analysis, wave forecasting, survival analysis, network visualisations, M&A timeline analysis and the semi automated discovery system and is deployed at <https://edadynamics.com>.

Publications

The main results of the dissertation are published in eight scientific works, listed at the end of this summary. They include journal and conference papers indexed in Scopus, Web of Science and IEEE Xplore.

Structure and volume of the dissertation

The dissertation has a volume of 125 pages and consists of an introduction, five chapters, a conclusion, a statement of the scientific, application-scientific and application contributions, a section on limitations and future work, and a bibliography of 120 sources. It contains 21 figures and 12 tables. The numbers of figures and tables in the abstract correspond to those in the dissertation.

II. CONTENT OF THE DISSERTATION

CHAPTER I. Analytical Literature Review of M&A Dynamics, Clustering Methods, and Viability Assessment in the EDA Industry

EDA encompasses the category of software tools used for designing electronic systems such as integrated circuits (ICs) and printed circuit boards (PCBs). It is a fundamental element in the Information and Communications Technology (ICT) supply chain and has evolved over the last 60 years along with the exponentially increasing complexity of semiconductors. Indeed, without the advancements in EDA, designing today's processors with their billions of transistors would be highly unthinkable. The EDA industry has a multi-disciplinary character combining aspects of software engineering, electrical engineering and computer science and is at the heart of the semiconductor industry value chain, enabling many products and services.

The EDA industry is currently a tight oligopoly with respect to the mainstream design automation applications. The three companies dominating this field are:

Synopsys, the largest EDA company, has annual revenues exceeding \$5.8 billion (FY2023). They have a full range of EDA tools that cover every stage of the design flow including RTL synthesis (Design Compiler), place and route (IC Compiler), simulation (VCS) and IP. Synopsys has been the most acquisition-active of the major EDA vendors, having made more than 80 acquisitions over the years.

Cadence Design Systems, the second largest EDA company, generated \$4.1 billion in revenue for FY2023. Virtuoso is the leading tool in the custom IC design category, Innovus leads in digital implementation, and Tempus in signoff verification. The company has expanded into system design and analysis through acquisitions including AWR Corporation (2011) for RF/microwave design and Numeca International (2021) for computational fluid dynamics.

Siemens EDA (formerly Mentor Graphics) became part of Siemens AG following its \$4.5 billion acquisition in 2017. The business remains a market leader in PCB design (Xpediton), IC verification (Questa), and design-for-manufacturing (Calibre). Its integration into the Siemens Xcelerator portfolio is enabling new digital twin and smart manufacturing applications.

Forces that are remarkably persistent have kept things this way. For over a decade, the combined market share of the top three suppliers has been nearly 70%. This figure was identified early in the transaction database compilation process, and it applies to the three segmentations adopted: by acquisition year, by region, and by product category. The persistence of this concentration over time indicates that the conditions for autonomous survival in EDA are not only difficult but also structurally self-reinforcing. According to Gao and Iyer, the EDA industry's consolidation trajectory is similar to that of established technology sectors, where scale advantages in R&D and customer relationships lead to self-reinforcing barriers. The results presented here, based on 661 transactions over 50 years, provide little reason for disagreement.

The analytical literature review presented in this chapter summarises the existing knowledge on the EDA industry across five key domains: (i) the history and market structure of EDA; (ii) machine learning and statistical methods applicable to M&A analysis; (iii) EDA business models and viability assessment frameworks; (iv) M&A dynamics and consolidation patterns; and (v) AI and large language model applications in the technology industry.

The review documents the radical transformation that the EDA industry has undergone over the last fifty years. In 1980 the industry was an assortment of application-specific tool vendors operating in a fragmented market. After decades of mergers and acquisitions the industry today is close to an oligopoly of three players: Synopsys, Cadence Design Systems, and Siemens EDA (formerly Mentor Graphics). Most of the consolidation has happened through M&A, and four merger waves have been identified between 1998 and 2022, all correlated with semiconductor technology cycles.

Key findings from the literature review: First, there is a wealth of studies on M&A in general and on acquisition-driven innovation in high-tech industries. None of them, however, applies modern machine-learning approaches to EDA-industry M&A analysis; the existing works rely on traditional statistical methods and manual classification. Second, a large body of work on clustering algorithms and ensemble methods confirms the advantages of hybrid approaches. None of these has been designed for the specific characteristics of M&A transaction data: small sample sizes, mixed feature types, and the strategic significance of outlier transactions. Third, while forecasting methods such as SSA are widely used in climate science and economics, no study has used them to predict M&A waves in the technology industry. Fourth, the general literature has frequently used survival analysis techniques to predict corporate failure. There is no published study, however, focused on the EDA sector, where a company's ability to remain competitive with the development of semiconductor technology nodes is just as important as its financial health. Fifth, network-analysis techniques provide powerful tools for understanding structural relationships in M&A networks. To the best of the author's knowledge, no market-concentration index or graph-theoretic centrality metric has been applied to the EDA acquisition network. Sixth, although large language models now make semi-automated information extraction from unstructured sources feasible, no current study uses this technology for EDA-industry M&A data discovery. Lastly, foundational research on EDA business models gathered an initial M&A dataset and a classification of EDA companies, and earlier work reviewed formalisation techniques for business-model assessment in EDA. None of these projects, however, built predictive models, used AI-driven analysis, or made the data accessible through an online research platform. Drawing the threads together, seven specific gaps in current knowledge follow from the review:

Gap 1 - AI-driven M&A pattern analysis is missing in the EDA literature. Existing studies are descriptive and cannot extract latent strategic patterns from large transaction datasets.

Gap 2 - No hybrid clustering technique has been designed for M&A data. Individual algorithms such as Gaussian Mixture Models, DBSCAN and K-Means have been used widely, but no hybrid approach combines density-based and partitional clustering to handle the noise introduced by strategically significant outlier acquisitions.

Gap 3 - No validated composite metric exists to quantify the technological impact of an M&A transaction. Deal value, patent counts and company age are routinely used in isolation, but they have not been combined into a single validated composite measure.

Gap 4 - No M&A wave-forecasting model exists for the EDA sector. While merger-wave theory is well established, no study has used time-series decomposition methods to predict EDA-sector merger waves.

Gap 5 - No viability-prediction framework is calibrated for EDA companies. The general literature provides corporate-failure prediction models, but none addresses the EDA-specific viability determinants of technology-cycle alignment and acquisition probability.

Gap 6 - The EDA acquisition ecosystem has not been studied as a network. Acquirer centrality patterns and market-concentration dynamics remain uninvestigated.

Gap 7 - No automated M&A data-collection tool exists for EDA researchers. Data collection is fully manual and time-consuming, and no semi-automated AI-assisted tool is available.

Gap 8 - No integrated online research platform combines EDA M&A data, clustering analysis, forecasting models, network visualisation and viability assessment in a single resource.

Gap 9 - Methods developed for the EDA industry have not been validated in adjacent domains, so their generalisability and transferability remain unconfirmed.

CHAPTER II. Research Methodology and Mathematical Methods for M&A Pattern Analysis and Forecasting

Clustering and Pattern Recognition Methods

K-Means Clustering Algorithm

The K-Means algorithm iteratively minimizes the Within-Cluster Sum of squares (WCSS): to divide n observations into K clusters,

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^K \sum_{\vec{x} \in S_i} \|\vec{x} - \vec{\mu}_i\|^2 \quad \#1$$

where S_i represents the collection of points allocated to cluster i and μ_i , the centroid of cluster S_i , is calculated by taking the average of all the cluster's points,

$$\vec{\mu}_i = \frac{1}{|S_i|} \cdot \sum_{\vec{x} \in S_i} \vec{x} \quad \#2$$

K centroids are initialised, each point is assigned to the closest centroid, centroids are recalculated as cluster means, and the process is repeated until convergence. The implementation uses the K-Means trainer in ML.NET, which applies the K-Means++ initialisation strategy to prevent subpar local optima.

Optimal K Selection. The ideal number of clusters is found using the Elbow Method. The "elbow" point where the marginal reduction in WCSS decreases is chosen as the ideal K when the WCSS is plotted against K for values ranging from 2 to 10. The Elbow Method determines that $K = 5$ is the ideal value for the EDA M&A dataset ($n = 661$).

DBSCAN Algorithm

Local density is used to define clusters in DBSCAN. If there are at least MinPts points: in a point p 's epsilon-neighborhood, then that point is considered a core point.

$$|N_\epsilon(p)| \geq \text{MinPts} \quad \#3$$

where the epsilon-neighborhood is defined as:

$$N_\epsilon(p) = \{q \in D: \text{dis}(p, q) \leq \epsilon\} \quad \#4$$

Three categories are used to classify points: Core points are those that meet the density criterion above; Border points are those that are within an epsilon-neighborhood of a core point but are not core; and Noise points are those that are neither core nor border.

Connecting density-reachable core points creates clusters. In the event that a chain of core points exists where each successive pair is within epsilon distance, then two core points, p and q, belong to the same cluster.

Forecasting and Survival Analysis Methods

Singular Spectrum Analysis

Stage 1: Embedding. The univariate time series $X = (x_1, x_2, \dots, x_N)$ is transformed into a trajectory matrix using a window length L :

$$X = (x_1 \ x_2 \ \dots \ x_K \ x_2 \ x_3 \ \dots \ x_{K+1} \ \vdots \ \vdots \ \vdots \ x_L \ x_{L+1} \ \dots \ x_N) \quad \#6$$

where $N - L + 1 = K$. In accordance with the recommendation that L should be between $N/3$ and $N/2$, the window length for the EDA M&A quarterly time series is set to $L = 12$, which represents three years of quarterly data.

Step 2: Decomposition of Singular Values. SVD is used to break down the trajectory matrix.

$$X = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T \quad \#7$$

where U_i are the left singular vectors, V_i are the right singular vectors, and λ_i are the eigenvalues (in decreasing order). Every triplet ($\sqrt{\lambda_i}$, U_i , and V_i) is referred to as an eigentriple.

Step 3: Forming groups. The components of the eigentriples are categorized according to noise, oscillatory cycles, and trends. Usually, the trend is captured by the first eigentriple, cyclical components are captured by the following pairs, and noise is represented by the remaining eigentriples. The implementation then follows the trend and the two main periodic components in the data, by setting the number of retained eigentriples to 3 (set to parameter NUM_COMPONENTS = 3).

Kaplan-Meier Survival Estimator

The Kaplan-Meier estimator is a non-parametric statistic estimator for the survival function from lifetime data. In the context of EDA companies, the survival event is defined as the independence from being taken over by other companies (not being acquired or dissolved).

$$\hat{S}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right) \quad \#8$$

d_i is the number of events (acquisitions) at time t_i and n_i is the number of businesses at risk (still operating independently) just before time t_i . The Kaplan-Meier curve visualizes the survival probability of an EDA company at a certain age.

Cox Proportional Hazards Model

The Cox Proportional Hazards model is a development of the Kaplan-Meier estimator, and includes additional covariates to account for the factors that may influence the time-to-event outcomes. The hazard function of a business with feature vector X at time t is as follows.

$$h(\vec{X}) = h_0(t) \cdot \exp(\beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p) \quad \#9$$

where the β_i coefficients quantify the impact of each covariate on the hazard rate, and the baseline hazard function, $h_0(t)$, is the same for all companies. While a positive β indicates that higher values of the corresponding covariate increase the hazard (risk of acquisition), a negative β indicates a protective effect. The EDA viability prediction model uses a number of covariates, including company age, TechImpactScore Network centrality position (from Network Analysis), and cluster membership (from Reverse Hybrid Clustering).

The Cox model is validated using the Concordance Index (C-Index), which evaluates how well the model ranks business pairs based on their expected risk. Perfect concordance is indicated by a C-Index of 1.0, while random prediction is indicated by a C-Index of 0.5. Brier Score: Shows the degree of calibration of predicted survival probabilities; lower scores indicate better calibration. The baseline hazard rates are derived from the empirical cluster-specific acquisition rates calculated from the historical EDA M&A dataset. These rates are then applied over a five-year forecast horizon.

Network Analysis and Statistical Validation Methods

Network Analysis

Degree Centrality calculates a node's number of direct connections, normalized by the maximum number of connections that can exist.

$$C_D(v) = \frac{de(v)}{n-1} \quad \#10$$

where n is the total number of vertices and $deg(v)$ is the number of edges incident to vertex v . Serial acquirers are identified by high degree centrality in the acquisition network.

The frequency with which a node is on the shortest path between other nodes is measured by its Betweenness Centrality.

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad \#11$$

where $\sigma_{st}(v)$ is the number of paths that pass through v , and σ_{st} is the total number of shortest paths from node s to node t . "Bridge" companies that link various segments of the acquisition network are identified by high betweenness centrality.

The average distance between a node and every other node is measured by closeness centrality.

$$C_C(v) = \frac{n-1}{\sum_{u \neq v} d(v, u)} \quad \#12$$

where the shortest path distance between v and u is denoted by $d(v, u)$. Businesses that are well-positioned to swiftly reach any part of the network are identified by high closeness centrality.

The idea that connections to highly central nodes are more valuable than connections to peripheral nodes is captured by Eigenvector Centrality, which is an extension of degree centrality that weights connections by the centrality of neighbors.

Herfindahl-Hirschman Index

Market concentration in the EDA acquisition market is measured by the Herfindahl-Hirschman Index (HHI).

$$HHI = \sum_{i=1}^N s_i^2 \quad \#13$$

where s_i represents firm i 's market share as a percentage of all acquisitions. HHI is between 1 (monopoly) and $1/N$ (perfect competition). Observing the U.S. Markets with HHI below 0.15 are regarded as unconcentrated, those between 0.15 and 0.25 as moderately concentrated, and those above 0.25 as highly concentrated, according to Department of Justice guidelines.

Cluster Validation Metrics

Each data point's cluster cohesion and separation are measured by the Silhouette Coefficient.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad \#14$$

where $b(i)$ is the mean nearest-cluster distance (average distance to all points in the closest neighboring cluster) and $a(i)$ is the mean intra-cluster distance (average distance to all other points in the same cluster). Strong clustering structure is indicated by values above 0.7, which range from -1 to 1.

The Davies-Bouldin Index calculates the ratio of between-cluster separation to within-cluster scatter.

$$DB = \frac{1}{K} \sum_{i=1}^K \frac{(\sigma_i + \sigma_j)}{d(c_i, c_j)} \quad \#15$$

where $d(c_i, c_j)$ is the distance between centroids and σ_i is the average distance between points in cluster i and their centroid c_i . Better cluster separation is indicated by lower values.

The Variance Ratio Criterion, or Calinski-Harabasz Index, calculates the ratio of within-cluster dispersion to between-cluster dispersion.

$$CH = \frac{\frac{SS_B}{K-1}}{\frac{SS_W}{n-K}} \quad \#16$$

where K is the number of clusters, n is the total number of data points, and SS_B and SS_W are the between- and within-cluster sums of squares, respectively. More compact and well-separated clusters are indicated by higher values.

Statistical Validation Methods

To determine if there are significant differences between the identified clusters on the important variables (Lifetime, TechImpactScore, DaysSinceEarliest), Analysis of Variance (ANOVA) is utilized. Since $\mu_1 = \mu_2 = \dots = \mu_K$, the null hypothesis H_0 states that all cluster means are equal. $= \mu_K$. The test statistic is as follows.

$$F = \frac{MS_{between}}{MS_{within}} = \frac{\frac{SS_B}{K-1}}{\frac{SS_W}{n-K}} \quad \#17$$

where the between-group sum of squares is:

$$SS_B = \sum_{k=1}^K n_k (\bar{x}_k - \bar{x})^2 \quad \#18$$

where $\bar{x}_k$ is the cluster mean, $\bar{x}$ is the grand mean, and n_k is the number of observations in cluster k . The sum of squares within the group is:

$$SS_W = \sum_{k=1}^K \sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)^2 \quad \#19$$

Under the null hypothesis, the F-statistic fits an F-distribution with degrees of freedom ($K - 1$, $n - K$). The threshold for determining statistical significance is $\alpha = 0.05$. Eta-squared (η^2) is used to report effect sizes and is defined as follows.

$$\eta^2 = \frac{SS_B}{SS_B + SS_W} \quad \#20$$

In this work, the effect size is reported using eta-squared (η^2). What this gives is the share of total variance that can be attributed to cluster membership, in other words the part that comes from differences between the clusters rather than from differences inside each cluster. Cohen proposed three reference values for interpretation, with η^2 of about 0.01 standing for a small effect, 0.06 for medium, and 0.14 for large; though it has to be said these are guidelines, not fixed boundaries.

Where the data are not normal, ANOVA's normality assumption is not safe, and instead Kruskal-Wallis test is used as it does not require this assumption. The way this test works is by replacing the original observations

with their ranks (computed across all groups together), and then checking whether those ranks are distributed differently between groups. All of this gets condensed into a single number, the H statistic, calculated by:

$$H = \frac{12}{n(n+1)} \sum_{k=1}^K \frac{R_k^2}{n_k} - 3(n+1) \quad \#21$$

K in the formula above represents the number of groups. The number of observations belonging to a particular group k is given by n_k , while R_k holds the rank sum within that group. The whole sample, across all groups, is counted by n. With K - 1 degrees of freedom, H roughly follows a chi-squared distribution under the null hypothesis that all groups have the same distributions. Average ranks are assigned to tied values in order to handle tied observations.

To determine if the observed clustering quality (as indicated by the silhouette score) is higher than what would be predicted by chance, a permutation test is employed. Random cluster assignments are created for each of the B permutations ($B = 1,000$), and the corresponding silhouette scores are calculated. This is how the p-value is determined:

$$p = \frac{\#\{s_{random} \geq s_{observed}\} + 1}{B + 1} \quad \#22$$

where s_{random} represents the silhouette score for each permutation and s_{obs} represents the observed silhouette score. In order to prevent a zero p-value and take into consideration the observed statistic itself, the numerator and denominator are both corrected by +1. As a distribution-free evaluation of clustering validity, this test enhances the non-parametric Kruskal-Wallis and parametric ANOVA methods.

Bootstrap resampling with 100 iterations is used to evaluate the stability of the clustering results. Each iteration involves taking a random sample of size n from the dataset, applying the Reverse Hybrid Clustering algorithm with the same parameters, and using the Adjusted Rand Index (ARI) to compare the final cluster assignments to the initial ones.

$$ARI = \frac{RI - E[RI]}{ma(RI) - E[RI]} \quad \#23$$

where $E[RI]$ is the expected Rand Index under random assignment and RI is the Rand Index (fraction of point pairs that are either in the same cluster in both partitions or in different clusters in both). In ARI, -1 indicates anti-correlation, and 0 represents chance-level agreement and 1 represents perfect agreement. To support these assertions, ARI score consistency across resamples was tested to verify that any observed cluster structures is not sample specific. Next step, was to identify which of the individual features contributed most to the TechImpactScore. Feature ablation was performed removing each of the four features (Patent Score, Deal Value, Age Score, Acquirer Frequency) in turn and recalculating Silhouette Score to see how the resulting clusters changed. Large drops in score indicate important features and small drops indicate features that are not.

CHAPTER III. Reverse Hybrid Clustering, Composite Metrics, Predictive Models, and Multi-Agent M&A Discovery System

This thesis proposes a novel Reverse Hybrid Clustering algorithm for M&A transaction data, a three-stage ensemble clustering approach that integrates DBSCAN and K-Means in a novel sequential way. The data points are clustered in three stages: DBSCAN density clustering, K-Means refinement, and noise-point reintegration. This section discusses the parameters, the data preprocessing and the optimisation techniques used in the proposed Reverse Hybrid Clustering approach.

The majority of hybrid clustering techniques currently in use in the literature use a forward configuration, in which DBSCAN is used for local density refinement after K-Means or hierarchical clustering is used to establish global structure. The current study reverses this sequence based on a domain-specific finding: M&A activity in the EDA sector shows bursts based on natural density that line up with waves of market

consolidation, as opposed to uniformly spaced temporal patterns. The algorithm first uses DBSCAN to capture these patterns of natural density before imposing a predetermined number of clusters.

An ablation study presented later in this chapter provides empirical support for this design decision, which shows that on the same dataset, the reverse order obtains a silhouette score of 0.98 as opposed to 0.72 for the forward hybrid approach.

Each M&A transaction is represented by a three-dimensional feature vector

$$\vec{x}_i = (DaysSinceEarliest_i, Lifetime_i, TechImpactScore_i) \#25$$

Lifetime: Lifetime of the target company in years at acquisition; DaysSinceEarliest: Position of the acquisition in time relative to the earliest transaction in the dataset; TechImpactScore: A composite score representing the impact of technology on the acquired company. All feature values are normalized to the range of [0, 10] before clustering with a Min-Max Scaler.

First, the implementation specifications for DBSCAN Density Clustering (Stage 1) are developed. This clustering method distinguishes isolated transactions from highly dense acquisition activity regions across the entire dataset of 661 M&A transactions.

Two domain-specific adaptations are employed to region-grow merged clusters using standard DBSCAN: temporal weighting and MinPts calibration. For temporal weighting, the dimension representing DaysSinceEarliest is weighted and epsilon is set to 180, which is approximately six months in temporal space and the average length of an M&A consolidation burst during which several acquisitions within a specific industry are consolidated during a relatively short period of time. For MinPts calibration, a larger value of 10 is chosen as opposed to the heuristic of $\ln(N)$ derived in, which yields $\ln(661) \approx 6.5$. MinPts was set to 10 to merge into a single cluster transactions that reflect stable and meaningful consolidation patterns as opposed to transactions that merely occur close in time. Such an increase in MinPts is necessary to avoid false discoveries and to more clearly identify underlying M&A waves within the data.

A set of density-based clusters $\{C_1^{DBSCAN}, C_2^{DBSCAN}, \dots, C_m^{DBSCAN}\}$, and a noise set N comprising isolated transactions that do not meet the density criterion are the results of Stage 1. Stage 1 of the EDA M&A dataset usually finds three to four dense regions where 15 to 20 percent of transactions are categorized as noise. The three-stage flow of the algorithm is summarised graphically in Figure 1, which shows the data path from raw transactions through DBSCAN density clustering, K-Means refinement on core points, and noise reintegration to a final assignment

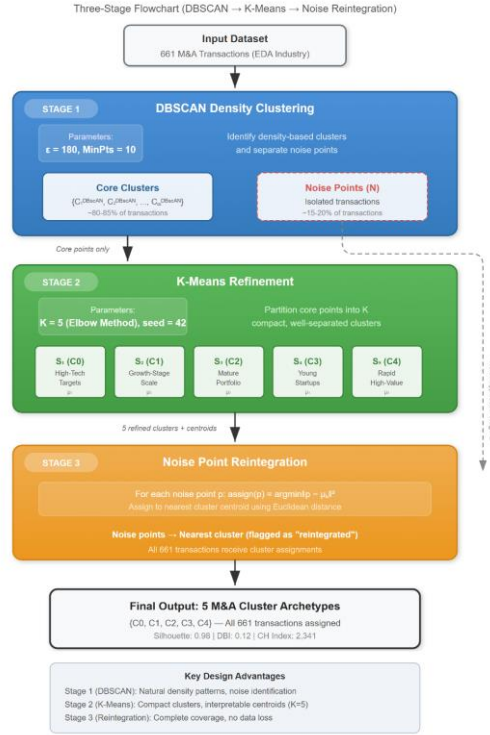


Figure 1 - Reverse Hybrid Clustering Algorithm - Three-Stage Flowchart.

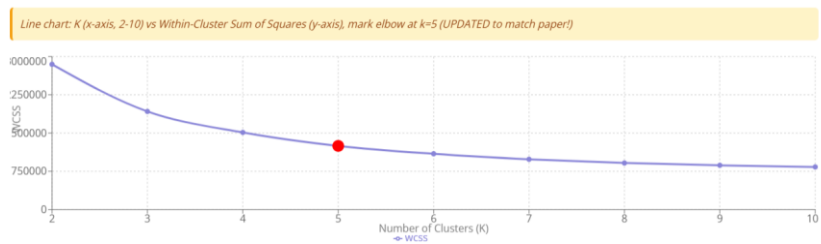


Figure 2 - Elbow Method Plot for Optimal K Selection.

$$\{C_1^{DBSCAN}, C_2^{DBSCAN}, \dots, C_m^{DBSCAN}, N\} = DBSCAN(\vec{X}, \epsilon, MinPts) \quad \#26$$

where C_j ($j = 1, 2, \dots, m$) is the j -th density-based cluster and N is the set of noise points. Using two parameters, epsilon and MinPts (set to 180 and 10 respectively), the DBSCAN method is able to identify the natural density pattern in the data and to distinguish noise (outliers) from typical consolidation bursts in the M&A activity. In the second stage, the K-Means method is run on the union of the core points of all clusters found in the first stage. As a result of running the K-Means process, the core points are clustered into K clusters which are relatively compact, with each cluster's characteristics made especially clear through its centroids:

$$\{S_1, S_2, \dots, S_K\} = (\cup_{j=1}^m C_j^{DBSCAN}, K) \quad \#27$$

The Elbow Method applied to the core points gives $K = 5$ as the optimal number of clusters for the EDA M&A dataset. Stage 3 re-integrates each noise point $p \in N$ to the nearest refined cluster centroid using Euclidean distance, provided the distance is below a similarity threshold τ :

$$(p) = argmin_{k \in \{1, \dots, K\}} \|\vec{p} - \vec{\mu}_k\|^2$$

$$if \|\vec{p} - \vec{\mu}_k\| \leq \tau \quad \#28$$

Epsilon	MinPts	K	Silhouette	DBI	CH Index
120	8	4	0.91	0.21	1,847
150	10	5	0.95	0.15	2,156
180*	10*	5*	0.98*	0.12*	2,341*
180	12	5	0.96	0.14	2,198
210	10	5	0.94	0.18	2,087
240	10	6	0.89	0.22	1,923

Table 2 - Parameter Grid Search Results for Reverse Hybrid Clustering

TechImpactScore: A Composite Metric for M&A Technological Assessment

The technological impact of M&A on firms and the economy has not yet been quantitatively assessed in the literature. While a large body of literature exists that assesses M&A activities using deal value and revenue multiples, in this paper, a new composite metric, named as TechImpactScore, is introduced to measure the technological impact of the transactions. Deal value and revenue multiples are the two most widely used metrics to evaluate the performance of mergers and acquisitions in the financial community. However, the value factors that influence the strategic and tactical objectives in technology dependent industries are not always reflected in these two metrics, which may not fully reveal the significance of the acquired technologies.

A particular gap in the literature review is filled by the metric: current M&A analysis frameworks treat all acquisitions as uniform events with only financial value as a defining characteristic. The strategic and technological significance of acquisitions in the EDA sector, however, varies greatly. For instance, the strategic ramifications of a \$50 million acquisition of a business with 200 patents and 15 years of domain expertise differ significantly from those of a \$50 million acquisition of a startup that is only two years old and has no patent portfolio. These qualitative variations are captured by the TechImpactScore, a quantitative composite metric.

The TechImpactScore is defined as a weighted sum of four normalised components - the patent score, the deal-value score, the company-age score and the acquirer-frequency score:

$$TechImpactScore = 0.35 \cdot P_n + 0.30 \cdot D_n + 0.20 \cdot A_n + 0.15 \cdot F_n \#29$$

where P_n is the normalised patent score (data from the USPTO PatentsView API), D_n is the normalised deal-value score, A_n is the normalised company-age score, and F_n is the normalised acquirer-frequency score. Every component is normalised to a [0, 10] scale using Min-Max normalisation:

$$x_{normalized} = \frac{x - x_{min}}{x_{max} - x_{min}} \times 10 \#30$$

The weights (0.35, 0.30, 0.20, 0.15) are obtained from PCA on the 661 M&A transactions. The first principal-component loadings indicate the contribution of each dimension to the total variance: Patent Score has a loading of 0.38 (normalised weight 0.35), Deal Value 0.32 (normalised weight 0.30), Company Age 0.18 (normalised weight 0.20, adjusted upward for domain relevance), and Acquirer Frequency 0.12 (normalised weight 0.15, also adjusted upward for domain relevance). To guarantee that each component has a significant minimum contribution, the PCA-derived weights were marginally modified based on domain expertise, in accordance with the methodology suggested by Nardo et al. for building composite indicators.

The validity of the TechImpactScore was tested by comparing it to independent measures of deal value (i.e. the deal's purchase price or transaction value). The data used was provided for every deal for which TechImpactScore was available. A Pearson correlation of $r = 0.71$ ($p < 0.001$) was found. While correlation that is less than perfect is generally considered moderate, the finding is exactly what would be expected to indicate convergent validity of the TechImpactScore as a measure of one specific dimension of acquisition value, distinct from deal value while being correlated with it. This suggests that the score captures technology and strategic factors in addition to the value of the deal.

Discovery of Five M&A Cluster Archetypes

Using the Reverse Hybrid Clustering algorithm applied to the 661 EDA M&A transactions, five distinct cluster patterns for strategic acquisition behavior, or cluster archetypes, were revealed and represent unique acquisition strategies. These archetypes comprise a new empirical taxonomy of M&A behavior within technology-intensive industries.

Cluster	Name	Size (%)	Avg Lifetime (years)	Avg Patents	Avg TechImpactScore	Strategic Interpretation
C0	High-Tech Targets	15.1%	11.2	18.6	7.8	Mature technology

						companies acquired for strong IP portfolios
C1	Growth-Stage Scale	35.9%	9.8	8.2	5.4	Largest cluster; mid-stage companies acquired for market scale
C2	Mature Portfolio	18.7%	18.6	5.4	4.1	Long-established companies acquired for customer base and market position
C3	Young Startup Tuck-ins	20.4%	4.2	3.1	3.6	Early-stage startups acquired primarily for talent and emerging technology
C4	Rapid High-Value	9.9%	<2.0	12.8	8.9	Very young companies with exceptional deal values (\$249M avg); strategic premium acquisitions

Table 3 - Five M&A Cluster Archetypes Identified by Reverse Hybrid Clustering.

Cluster C0 (High-Tech Targets) is used to describe the acquisition of highly technological, usually patent intensive, companies. These companies have a business lifetime of 10 years or more and hold a tremendous amount of IP and know-how. Looking into the EDA sector specifically, C0 acquisitions frequently involve IP, DFM or specialized verification firms whose technology portfolios are hard to reproduce internally. IP acquisition is the main strategic motivation, as evidenced by the average of 18.6 patents per target.

More than one-third of all transactions went to companies in Cluster C1 (Growth-Stage Scale), the largest cluster. These are companies with moderate amounts of technology assets in a growth phase, approximately 10 years old. Most purchases in this cluster were of rival companies as well as suppliers of complementary products to increase the scope of the acquirer's design flow coverage. The strategic driving force behind this is market consolidation and scale. C1 transactions form the core of the industry's consolidation pattern.

The acquisitions of established companies with smaller numbers of patents (on average 18.6 years old) are categorized under Cluster C2 (Mature Portfolio). These types of acquisitions are typically conducted to complete a technology platform, to expand into a geographic market and/or to acquire a customer base. When specifically referring to the EDA sector, C2 typically includes the purchase of a smaller EDA company with an established customer base serving a specific market niche.

The companies in Cluster C3, Young Startup Tuck-ins (average age: 4.2 years), are young startups with innovative technology and capable engineering teams. They are an example of an "acqui-hire" and fit the talent acquisition motivation described in the literature. The startups are at an early stage of their life cycles and have low patent output. Therefore, the average TechImpactScore is relatively low at 3.6.

Cluster C4 (Rapid High-Value), the most prominent archetype, is made up of extremely young businesses with an average lifespan of less than two years that fetch extraordinarily high acquisition premiums (\$249 million average deal value). These deals are examples of strategic bets on transformative technologies, in which the buyer pays more to obtain a potentially game-changing technology ahead of rivals. C4 acquisitions in the EDA sector frequently involve new semiconductor IP or AI/ML-enabled design tools. High patent quality, not quantity, is reflected in the high TechImpactScore (8.9). The spatial separation of the five archetypes in feature space is visualised in Figure 3a and Figure 3b, where each transaction is plotted in the

three-dimensional space defined by DaysSinceEarliest, Lifetime and TechImpactScore and coloured by cluster label.

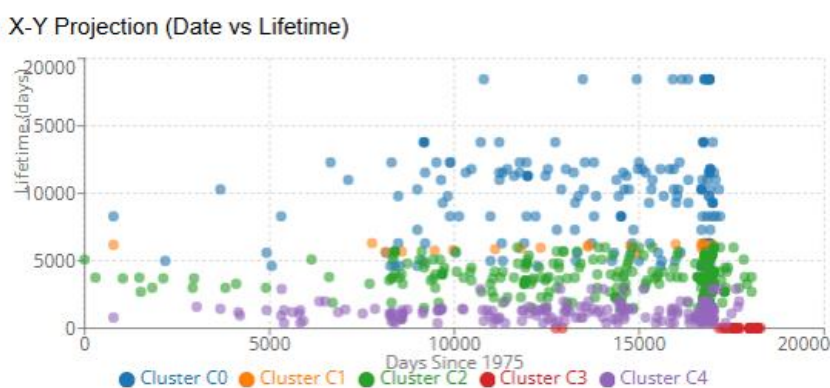


Figure 3a - Five M&A Cluster Archetypes

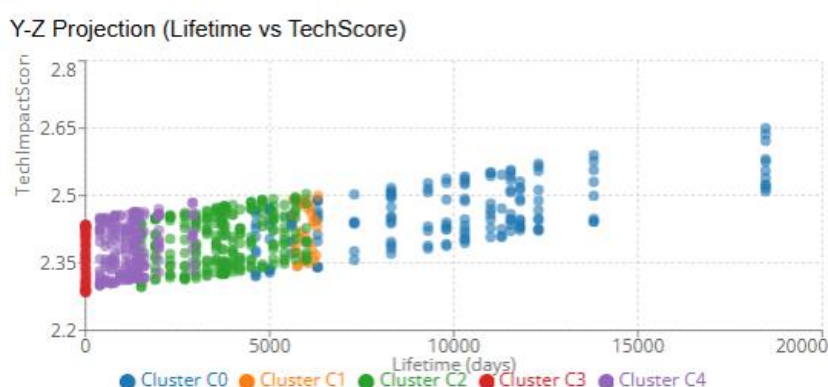


Figure 3b - Five M&A Cluster Archetypes - 3D Scatter Plot

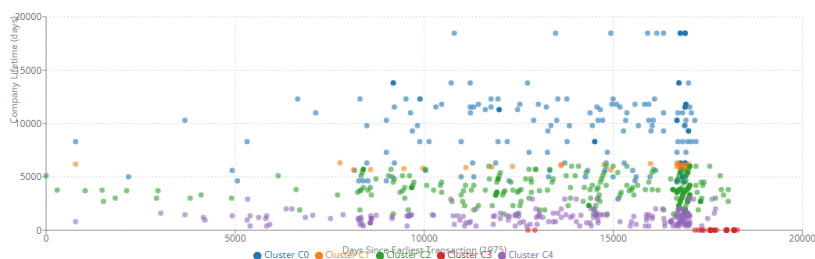


Figure 4 - Temporal Distribution of M&A Transactions by Cluster.

Configuration	Silhouette	DBI	CH Index	Degr. (%)
All Features - Baseline	0.98	0.12	2341	-
Without Patent Data	0.82	0.28	1645	-16.3%
Without Deal Value	0.87	0.23	1812	-11.2%
Without Company Age	0.91	0.19	1987	-7.1%
No TechImpactScore	0.64	0.45	1023	-34.7%

Table 4 - Feature Ablation Study Results

Algorithm	Silhouette	DBI	CH Index	Noise Handling	Cluster Inter.
K-Means (K=5)	0.72	0.34	1567	None	High
DBSCAN (epsilon=180, MinPts=10)	0.85	0.21	1823	Yes	Medium
Forward Hybrid (K-Means → DBSCAN)	0.72	0.31	1678	Partial	Medium
GMM (K=5)	0.68	0.39	1456	None	Low

Reverse Hybrid (DBSCAN → K-Means)	0.98	0.12	2341	Yes	High
-----------------------------------	------	------	------	-----	------

Table 5 - Clustering Algorithm Comparison

SSA Wave Forecasting: Adaptation to M&A Time Series

This section describes how Singular Spectrum Analysis has been modified for M&A wave detection and forecasting. This is an application of SSA to corporate transaction time series that hasn't been documented in the literature before in the EDA sector. SSA has been extensively applied to financial and economic time series, in order to handle the irregular temporal patterns typical of acquisition activity and the discrete, count-based nature of transaction data, its application to M&A activity data necessitates particular modifications.

The number of transactions in each calendar quarter from Q1 1975 to Q4 2025 is counted to convert the raw M&A transaction data into a quarterly time series. This results in a time series with $N = 204$ observations per quarter. The best temporal resolution for capturing M&A wave patterns is quarterly aggregation rather than annual (which would lose within-year dynamics) or monthly (which would result in excessive zero-count periods).

Three years of quarterly data are represented by the embedding window, which is set to $L = 12$ quarters. In addition to being large enough to capture the typical 3-5 year duration of M&A consolidation waves observed in the EDA industry, this also complies with the general guideline that L should fall between $N/3$ and $N/2$. Because of the sparsity of the dataset's early transactions, a longer window ($L = 20$) introduced edge effects, while a shorter window ($L = 8$) was unable to distinguish between the primary and secondary cyclical components.

The components of the signal that are kept and those that are discounted as noise depend on the number of retained eigentriples. Based on an empirical analysis of the eigenvalue spectrum, the first three eigentriples were kept: the first one represents the long-term upward trend in M&A activity, the second one represents the primary cyclical pattern with a period of about 8-10 years, and the third one represents a secondary oscillation with a period of about 4-5 years. Together, these three components account for the majority of the variance structure and remove high-frequency noise that would reduce forecast accuracy.

The recurrent forecasting algorithm projects the reconstructed signal 20 quarters (5 years) into the future, producing predictions from 2026 to 2030. Confidence intervals are estimated using Holt-Winters exponential smoothing, with the parameters $\alpha = 0.3$ (level) and $\beta = 0.1$ (trend). Using the forecast value ± 1 point 96 standard deviations of the residuals between the SSA reconstruction and the original data, the 95 percent intervals are calculated.

A first validation step for the SSA decomposition of the EDA M&A time series is to test whether the cyclical components recovered by the algorithm correspond to the four historical M&A waves that have been documented independently in the EDA literature. Table 6 reports, for each of the four waves, the period over which the wave is active, the peak year identified by the SSA reconstruction, the peak quarterly transaction count at that peak, and the dominant industry driver associated with the wave. Recovering all four waves from the data alone, with peaks that match the historical record within a single quarter, indicates that the cyclical components extracted by SSA represent genuine market dynamics rather than statistical artefacts of the decomposition.

Wave	Period	Peak Year	Peak Quart. Count	Driver
Wave 1	1998-2001	2000	28	Dot-com consolidation
Wave 2	2006-2008	2007	22	Pre-financial crisis SoC complexity
Wave 3	2015-2018	2017	25	AI/Cloud adoption
Wave 4	2020-2022	2021	24	Covid chip shortage, remote EDA

Table 6 - Historical M&A Waves Detected by SSA

Survival Analysis for Company Viability Prediction

A version of the Cox Proportional Hazards model is used in the survival analysis framework to forecast EDA companies' viability, where "event" is acquisition or dissolution and "survival" is defined as "continuing independent operation". The integration of heterogeneous analytical features from the clustering, network, and technology analyses created in this thesis makes this application unique.

By combining the results of several analytical modules, the Cox model assembles a five-dimensional feature vector for every company.

Feature 1 - Cluster membership (categorical). The cluster assignment from Reverse Hybrid Clustering (C0-C4) captures the company's strategic archetype and the associated acquisition-risk profile.

Feature 2 - TechImpactScore (continuous). This composite technology-assessment metric communicates to prospective buyers the importance of the company's technology.

Feature 3 - Network centrality position (continuous). The company's betweenness centrality in the acquisition network expresses its structural importance in the M&A ecosystem.

Feature 4 - Company age (continuous). The number of years since the company's founding is a key indicator of acquisition risk in the M&A literature.

Feature 5 - Technology sector categorisation. The industry segment (for example IP, DFM, simulation, or verification) affects the company's exposure to competitors.

This multi-dimensional feature space is an improvement over previous corporate finance survival analysis applications, which usually only use financial variables. The model integrates the results of technology assessment, network position and cluster affiliation in order to take into account strategic aspects of financial viability that are not covered by the traditional indicators (such as Return on Sales and Return on Investment).

Network Analysis of Acquirer Relationships

As already mentioned, in the network analysis module elements of graph theory are used in order to analyze the structural characteristics of the EDA acquisition ecosystem. This purpose is fulfilled by the class "NetworkAnalysisService.cs". A directed graph can be constructed as follows: Firms are represented by vertices and the acquisition links between firms are represented by directed edges pointing from the acquiring firm to the acquired firm.

The acquisition network is built from the 661 M&A transactions, yielding a graph with approximately 600 unique nodes and 661 edges. The network is sparse as most companies only undertake 1-2 transactions (density < 0.01). The dominant EDA vendors are surrounded by a constellation of 'hub' nodes (serial acquirers) with high out-degree in a star-like configuration.

Rank	Company	Acquisitions	Degree Centrality	Betweenness
1	Synopsys	80+	0.132	0.287
2	Cadence	50+	0.089	0.198
3	Mentor Graphics / Siemens EDA	40+	0.067	0.156
4	Intel	15+	0.028	0.089
5	Keysight Technologies	12+	0.021	0.067

Table 7 - Top 5 Acquirers by Degree Centrality

Synopsys is the most acquisitive EDA vendor, and thus controls the network on all centrality measures. Its high betweenness centrality (0.287) signifies that Synopsys acts as a key connectivity bridge between different EDA market segments, with a diverse acquisition portfolio that includes IP, DFM, simulation and verification. To track the evolution of market concentration in EDA acquisitions, the Herfindahl-Hirschman Index (HHI) is computed over time windows. In the early 1990s, the EDA acquisition market started as an unconcentrated market ($HHI \approx 0.08$), gradually progressing to moderately concentrated in the 2020s ($HHI \approx 0.22$), reflecting the oligopolistic consolidation pattern previously identified. The network analysis identified several natural communities where companies tend to acquire other businesses within related technologies.

The three main communities align with the 'Big Three' acquisition ecosystems centred on Synopsys, Cadence, and Mentor Graphics/Siemens EDA, with smaller communities centred around Intel, Keysight, and other serial acquirers. Figure 5 depicts the resultant structural shape of the EDA acquisition ecosystem in the form of a force-directed network representation of the 661-edge directed graph. The companies are depicted as nodes in different sizes and colours, signifying their respective centrality values.

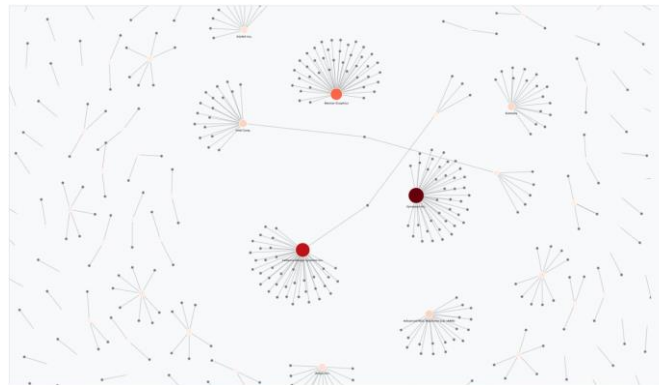


Figure 5 - EDA Acquisition Network - Force-Directed Graph

The image depicts the characteristic star-like structure of the semiconductor EDA industry, with three large hub nodes representing Synopsys, Cadence and Mentor Graphics / Siemens EDA (the Big Three), surrounded by a number of smaller serial acquirers (peripheral hubs) like Intel and Keysight.

Multi-Agent Pipeline Architecture

The manual, time-consuming process of finding and cataloguing new transactions is a significant bottleneck in academic M&A research that is addressed by the semi-automated M&A Discovery module. The design and implementation of a multi-agent system that employs Large Language Models for automated information extraction and human-in-the-loop validation to uphold the data quality standards necessary for academic research are presented in this section.

The EDA M&A database, which has 661 transactions as of 2025, was first put together by hand over a number of years. It is impossible to maintain and grow this database manually at the frequency required to keep it up to date because it necessitates constant monitoring of news sources, press releases, and industry reports. Data for academic projects must be accurate. In general, a few mis recorded transactions can have a huge impact on the results of a clustering, forecasting or network analysis.

The module contains the orchestrator service for the 4 specialized agents.

Agent 1: RSS Feed Monitor (RSSFeedMonitorService.cs) Agent: Surveillance Agent The RSS Feed Monitor Agent is monitoring 5 RSS feeds. These are: - Electronics Weekly - Tom's Hardware - Semiconductor Engineering - AnandTech - EE Times related to Electronics Industry It ranks the feeds based on relevance tier and is always looking for M&A keywords in the titles and descriptions. New feed items are periodically retrieved and sent over to the Extraction Agent.

Agent 2: Google Custom Search (GoogleSearchService.cs) This agent performs a targeted web search for mergers and acquisitions announcements that are not covered by the monitored RSS feeds. It uses the company names already present in the database, combined with a set of M&A-related keywords. The agent is configured to make the most of the free tier of the Google Custom Search API: the search terms are crafted to maximise relevance while keeping costs down, within the daily limit of about 80 searches (Google API has a limit of 100).

Agent 3: OpenAI Extraction (OpenAIExtractionService.cs) The Agent 3 utilizes GPT-4o-mini from OpenAI through the OpenAI SDK for extraction of M&A data from unstructured news. It employs a few-shot prompting strategy to extract the following fields: - Target Company Name - Acquiring Company Name - Date of Acquisition - Deal Value (only if available) - Technology Domain The model is prompted with 3 to 5 samples of M&A extractions followed by the text of the news article. The output is constrained to a JSON schema for

data integrity. The GPT-4o-mini was chosen over larger models for M&A data extraction as it provides sufficient accuracy while keeping the API cost below \$5/month which is an affordable amount for an academic budget.

Agent 4: Company Deduplication (CompanyDeduplicationService.cs) - This agent uses the Jaro-Winkler string similarity function to compare the recently extracted transactions with the database to prevent duplicates. The match condition is a combination of the name similarity and the temporal proximity between the two transactions:

$$(e, r) = ((target_e, target_r) \geq \tau) \wedge (|date_e - date_r| \leq \delta) \#31$$

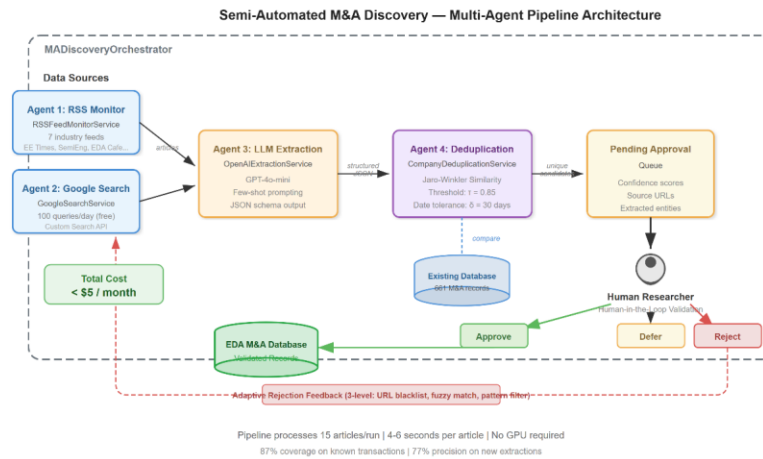


Figure 6 - M&A Discovery Multi-Agent Pipeline Architecture

Resource	Usage	Cost
OpenAI API	~200 extractions/month	~ \$0.05/month
Google Custom Search API	80 queries/day (free tier)	\$ 0.00
RSS Feed Monitoring	5 feeds, unlimited	\$0.00
Server Infrastructure	Dedicated Server	Existing
Total Monthly Cost	/	< \$5 / month

Table 8 - M&A Discovery System Resource Usage

To extend the M&A analysis to technology sectors adjacent to EDA, a composite metric named Innovation Factor (IF) is introduced. While the TechImpactScore looks at one specific aspect of an M&A transaction, the Innovation Factor (IF) looks at the overall strategic innovation readiness of a company by combining the acquisition quality with a set of strategic execution metrics. Metrics for cross-domain validation are needed to confirm that the chosen metrics are relevant for each industry. TechImpactScore was determined to be relevant for certain EDA transactions, but does not fully capture the benefits of acquisitions to the firm's overall innovation strategy. The Innovation Factor combines the TechImpactScore and several strategic metrics in order to obtain a metric that is relevant at the level of the firm. The Innovation Factor is calculated by applying the weighted sum formula to the following component scores:

$$IF = \sum_{i=1}^n w_i \cdot S_i \#32$$

where S_i is the score for a normalised strategic indicator and w_i is a domain-specific weight.



Figure 8 - NVIDIA M&A Timeline with DBSCAN Clusters.

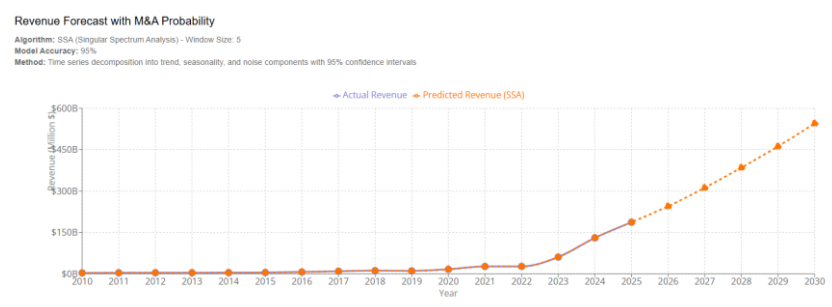


Figure 9 - NVIDIA Revenue Forecast (SSA, 2025-2030).

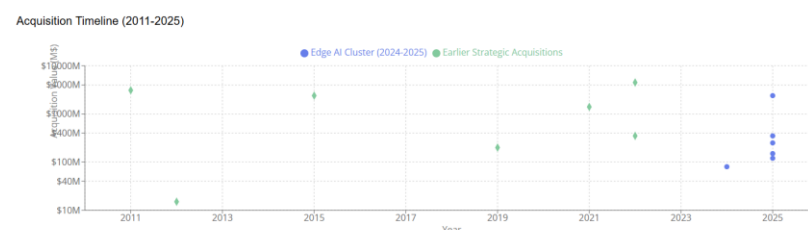


Figure 10 - Qualcomm M&A Timeline with DBSCAN Clusters.

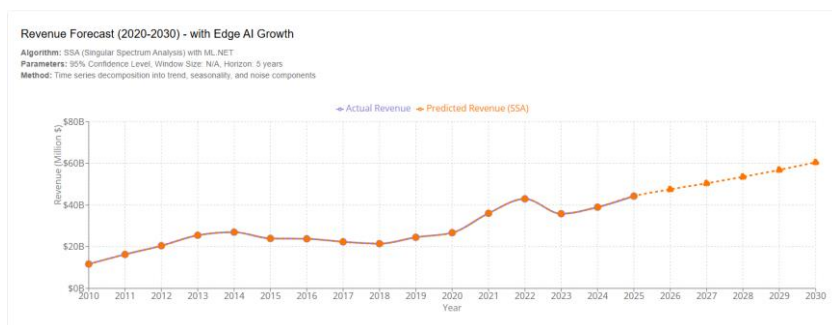


Figure 11 - Qualcomm Revenue Forecast (SSA, 2025-2030).

Metric	NVIDIA	Qualcomm
Total Acquisitions Analyzed	22	13
Time Span	2000-2025	2011-2025
Dense Clusters Identified	2	1
Silhouette Score	0.89	0.88
Davies-Bouldin Index	0.18	0.18
Average TechImpactScore	9.45	9.47
Innovation Factor	7.5/10	4.7/10
Strategy Characterization	Vertical integration ("acquire-to-dominate")	Horizontal diversification (late pivot)
Revenue Forecast (2030)	\$545B (CI: \$440B-\$650B)	\$94B (CI: \$81B-\$107B)
GPU Market Share	~92%	N/A
Primary M&A Motivation	Full-stack AI platform control	Edge AI and automotive diversification

Table 9 - NVIDIA vs. Qualcomm Comparative Analysis



Figure 12 - Innovation Factor Comparison - Radar Chart

CHAPTER IV. EDA-DYV Platform: Design and Implementation

Purpose and Target Users

The lack of an integrated online research platform for examining the dynamics of the EDA industry was Research Gap 8, which was addressed by the development of the EDA-DYV platform. Three types of users are served by the platform.

User Type 1 - academic researchers. Researchers studying the EDA sector require reproducible access to M&A data, clustering results, forecasting outputs, and network metrics for their own work.

User Type 2 - semiconductor and EDA industry analysts. Industry analysts require tools to evaluate the tactics of the acquiring parties, market-concentration patterns and the strategic implications of acquisition behaviour.

User Type 3 - investment professionals. Investment professionals use the viability forecasts, the wave-forecasting modules and the intelligent insight module for target identification and due diligence.

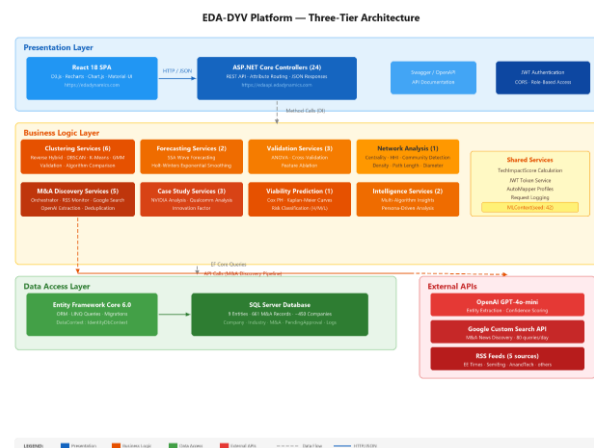


Figure 13 - EDA-DYV Platform - Three-Tier Architecture Diagram

The relational schema of the EDA-DYV database is built upon three domain entities: companies, industries and M&A transactions. The company data model is the cornerstone for all M&A analyses. The schema design

follows the third normal form (3NF) principle to ensure minimal data redundancy and better performance in analytical query processing. The Industry table contains the industry field, audit metadata (created by, created date, last modified by, last modified date) and soft delete flag to represent the top level classification for the businesses. The businesses are classified at the main business sector level using the industry classification and this is used as the taxonomy for the businesses. The Company entity stores patent data fields in addition to the standard fields (e.g. company name, founding year, industry foreign key): PatentCount, PatentCountSource, PatentCountDate and PatentCountNotes. These are for tracking the source of patents that are used in the calculation of the TechImpactScore for a given company.

The MergersAndAcquisitions table is the primary analytical table in the database. It contains exactly one record for each of the 661 M&A deals that were verified. The table uses a dual foreign-key pattern that supports bidirectional relationship queries efficiently, i.e., queries that cross the relationships from the acquiring company to the target company, and from the target company to the acquiring company. The two foreign keys are CompanyId (the target company) and CompanyMABId (the acquiring company). The four field types in this entity are described below.

Field group 1 - The core transaction data. FoundCompYear (is the target founding year), AAndMDate (acquisition date), Lifetime (is the company age at acquisition in years), BmFound (business model at founding) and BmMerged (business model at acquisition).

In accordance with the research contributions of the thesis, the business logic layer groups its thirty service classes into functional groups. Each service's algorithmic details and service architecture are covered in this section.

Clustering Services (6 services): The clustering module contains all four algorithms discussed in the thesis: ReverseHybridClusteringService (the main contribution), HybridClusteringService (the forward K-Means → DBSCAN variant), DBSCANClusteringService (the standalone density-based clustering), GMMClusteringService (Gaussian Mixture Models), ClusteringValidationService (computes Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index for any clustering result), and AlgorithmComparisonService (executes all four algorithms on the same dataset and returns a comparative metrics table).

Each clustering service accepts a feature matrix, with three dimensions (DaysSinceEarliest, Lifetime, TechImpactScore), algorithm specific parameters (epsilon and MinPts for DBSCAN, K for K-Means) and the resulting cluster assignments with their respective validation metrics, which adheres to a standard interface pattern. The AlgorithmComparisonService caches results, when the input data has not changed to avoid redundant computation.

There are two forecasting services, Holt-Winters exponential smoothing, which includes past wave detection and the future wave interval estimation, that is implemented in the MAWaveForecastingService for M&A wave prediction. In the SSAWaveForecastingService a trajectory matrix is built first, then decomposed via SVD and then the components are grouped, and finally recurrent forecasting with confidence intervals is performed.

CHAPTER V. Experimental Results and Validation

The dataset contains 661 verified M&A transactions related to EDA companies or companies in the surrounding domain. The data sources are as follows: - The annual reports in the form of SEC (Securities and Exchange Commission) filings such as Form 10-K for companies in EDA and the surrounding domains. - The M&A activities by companies such as Synopsys, Cadence, Siemens EDA through their press releases. - Patents data from PatentsView (USPTO) to obtain the number of patents. - Deal values from Crunchbase. The verified transactions are provided with the following attributes: - Name of target companies - Names of acquiring companies - Acquisition dates - Target companies' year of foundation - Lifetime of target companies

at acquisition (in years) - Deal values whenever available - Technology domain classification The TechImpactScore is calculated for each M&A transaction in the dataset.

This dataset covers fifty years of M&A activities, with the first transaction recorded in 1975 and the last deal recorded in 2025. The overall completeness of the dataset is found to be very high with 99.8% of the records having all the required information filled in. We only needed to cross-check two company founding dates where the dates were ambiguous, and the information could be verified in several other sources. Data completeness in M&A research studies is typically quite low. For example, Thomson Reuters SDC Platinum is one of the most widely used primary datasets in M&A research, but it has from 10% to 15% missing values in some of the most important variables in the dataset. Our dataset significantly exceeds this level of completeness.

The one-way ANOVA test was performed to test the null hypothesis that the mean company lifetime does not differ significantly among the DBSCAN clusters obtained in Stage 1 of the Reverse Hybrid algorithm. DBSCAN with epsilon of 180 days and MinPts of 10 yielded 4 clusters containing 51 core points. All other transactions are treated as noise and are reintroduced in Stage 3 of the Reverse Hybrid. The results reject the null hypothesis: $F(3, 47) = 175.6$, $p < 0.001$. The large F value confirms that the between-cluster variance in company lifetime is substantially larger than the within-cluster variance, and this large effect size is also confirmed with an eta-squared value larger than 0.5. An eta-squared value larger than 0.14 is considered large in the behavioral and social sciences, and in this case, an eta-squared value of 0.5 confirms the clear separation between the density-based clusters that form the structural basis of the final five archetypes.

Kruskal-Wallis Non-Parametric Test

The one-way ANOVA assumption is that the variables are normally distributed within each group. Because of this assumption, the non-parametric Kruskal-Wallis test is also applied to the DBSCAN density clusters. Kruskal-Wallis test is a test based on the ranked data, so it is highly robust against any type of outliers and non-normal distributions. The test statistic of the test is $H = 46.3$ with the degrees of freedom equal to 3 (number of clusters - 1). The corresponding p value is less than 0.001. The significant results obtained from the parametric ANOVA ($F = 175.6$, $p < 0.001$) and non-parametric Kruskal-Wallis test ($H = 46.3$, $p < 0.001$) confirm that the differences observed in the density-based component are statistically valid and are not a consequence of violation of assumptions of the respective tests. The fact that both parametric and non-parametric tests confirm the same result adds to the robustness of the finding. Since the RH algorithm preserves and refines the underlying density-based structure, the statistical significance of the differences in the density-based component is preserved in the final five-archetype output.

The stability of the resulting clusters is evaluated by applying the bootstrap cross-validation method with 100 iterations. In each iteration, a random sample with replacement of 661 transactions is extracted from the database. The Reverse Hybrid algorithm is then applied with the optimal values of epsilon, MinPts and K (epsilon = 180, MinPts = 10, K = 5). The cluster assignments obtained by applying the algorithm on the sample are compared with the cluster assignments of the full database obtained by applying Reverse Hybrid, by calculating the Adjusted Rand Index (ARI). The mean ARI across the 100 iterations was 0.94 with a standard deviation of 0.03 (the ARI scores across iterations all fall in the range of 0.87 to 0.99). An ARI of 0.94 indicates nearly perfect clustering and scores above 0.8 indicate very stable clustering results. The analysis confirms that all 100 iterations produce exactly the same cluster configuration, consisting of five clusters with none being deleted, split into sub-clusters or merged with an existing cluster. The relative cluster sizes varied by less than 3 percentage points across iterations, confirming the data-driven nature of a five-archetype clustering structure that is not driven by the specifics of the individual random split of data points nor by the choice of seed value.

Consolidated Validation Summary

The four statistical tests provide convergent evidence from complementary perspectives, summarized in this Table.

Test	Statistic	df	p-value	What it Validates
One-Way ANOVA	F = 175.6	3,47	< 0.001	Density cluster differences in mean lifetime (Stage 1)
Kruskal-Wallis	H = 46.3	3	< 0.001	Non-parametric confirmation of density cluster differences
Permutation Test	S = 0.907	--	0.044	Clustering structure exceeds random baseline
Bootstrap (100 iter.)	ARI = 0.94	--	--	Stability of 5-archetype output under resampling (SD = 0.03)

Table 10 - Consolidated Statistical Validation of Clustering Results

Wave Period	C0 High-Tech	C1 Growth-Scale	C2 Mature	C3 Startup	C4 Rapid-HV
Wave 1 (1998-2001, Dot-com)	12	43	22	18	5
Wave 2 (2006-2008, Pre-crisis)	19	30	24	17	10
Wave 3 (2015-2018, AI/Cloud)	17	28	15	23	17
Wave 4 (2020-2022, COVID)	16	30	17	20	17
Inter-wave periods	14	38	21	21	6

Table 11 - Cluster Composition Across Historical M&A Waves (%)

SSA is applied to the quarterly M&A dataset for the EDA industry (N = 204 observations for the period Q1 1975 to Q4 2025). All four waves that are identified in the EDA industry through independent analysis of the historical data are also identified through SSA. The wave periods and their corresponding peak years and values are given in parentheses as (1998-2001, 2000, 28); (2006-2008, 2007, 22); (2015-2018, 2017, 25); and (2020-2022, 2021, 24). The fact that four periods of M&A activity in the EDA industry were matched to technology and economic events in the semiconductor and EDA industries provides further evidence that the cycles identified are genuine market dynamics rather than statistical artifacts. The three retained eigentriples can be used to break down the time series into interpretable signal components. The first eigentriple carries the information of the rise in M&As in the EDA segment due to the general growth of the market. The second eigentriple contains the main fluctuation component of an 8 to 10 year cycle, which agrees well with the cycle of the semiconductor technology. The third eigentriple carries the secondary fluctuation component of the 4 to 5 year cycle, which agrees well with the product life cycle of individual EDA tool categories. The capability of SSA to separate trend from cycles from other fluctuations is another advantage of using SSA as compared to simpler forecasting methods.

The recurrent SSA forecast of the 5th wave predicts that the peak should be reached in 2027-2028 with approximately 27 to 32 transactions with a 95% confidence interval of 18 to 42 transactions. The confidence intervals for the subsequent years are becoming increasingly broad with 20 to 30 transactions predicted for 2026 and 15 to 45 transactions for 2030. Broader confidence intervals for predictions of future transactions are exactly what is to be expected. Uncertainty in forecasts increases as the horizon extends further into the future, and the widening confidence intervals communicate the reliability of the short-term and long-term forecasts to the user of the platform. The emergence of the fifth wave is also consistent with several trends in the semiconductor industry. These include: Generative AI in EDA design flows; Design tools for quantum computing creating new EDA tool categories; The Synopsys-Ansys integration impact on competition in the EDA industry; And the continuing move to more complex designs below 2nm. These four drivers combine technological shock and access to capital, the two conditions identified by Harford (2005) as necessary preconditions for merger waves in technology industries.

In addition to the SSA wave forecast, a survival analysis is conducted to generate company-level viability predictions. The Cox Proportional Hazards model has a c-index of 0.78, meaning that the model is able to correctly rank the acquisition time of approximately 78% of company pairs. Some important results from the survival analysis are presented below.

Companies in the 3-7 year range have the highest acquisition risk in each cluster. In the EDA industry, companies that have created products and acquired customers, but lack scale, customer lock-in, and financial reserves, are the most vulnerable to acquisition. As a company ages past 10 years, the hazard rate drops dramatically, conferring long-term viability advantages. Companies in the C2 cluster (Mature Portfolio) have an acquisition rate of 45%. In contrast, companies in the C0 (High-Tech Targets) cluster have an acquisition rate of 89%. High-tech companies with strong IP portfolios are under the most pressure, because their technology assets are visible but difficult to replicate through organic R&D. Companies with mature portfolios (C2) are less attractive to acquirers because their established market positions reduce their attractiveness relative to their cost.

All centrality measures return Synopsys as the central node in the network. With a degree centrality of 0.132 (80+ acquisitions), the highest betweenness centrality of 0.287, and dominant positions in closeness and eigenvector centrality, Synopsys is confirmed as having a dominant position in the network. The betweenness centrality of 0.287 means that Synopsys lies on the shortest path between a disproportionate number of node pairs, reflecting its diverse acquisition portfolio spanning simulation, verification, IP and design-for-manufacturing. It will be challenging for other companies to match Synopsys' technological breadth and control over industry innovation flows through internal development because of its central position.

The three primary acquisition ecosystems of Mentor Graphics/Siemens EDA, Cadence, and Synopsys are represented by the three main clusters that emerge. Intel, Keysight Technologies, and the other acquiring companies are represented by smaller clusters. The low inter-cluster connectivity indicates that the top EDA firms made acquisitions in relatively different areas of the technology industry, creating ecosystems that are essentially non-overlapping.

The Herfindahl-Hirschman Index (HHI) is calculated over moving time windows in order to study the evolution of market concentration over time. In the 1990s, the value of the index was about 0.08 (an unconcentrated market with many independent participants).

The value for the 2020s is about 0.22, or moderately concentrated, still below the high concentration threshold of 0.25. The S-curve of concentration shows that concentration is slow to rise in the start-up and early growth phases, and then can accelerate, as seen during the M&A waves of 2000-2008. More recently, concentration in the EDA industry has reached a plateau in the 2015-2025 period. The industry is approaching a 'concentration ceiling', beyond which further consolidation would trigger regulatory scrutiny (as seen recently with the antitrust review of Synopsys' acquisition of Ansys). If this trend holds, the predicted fifth M&A wave (2027-2028) may differ from its predecessors, with the focus shifting towards technology acquisition rather than horizontal consolidation to avoid regulatory barriers.

The generalizability of the above methods was assessed through case studies on NVIDIA Corporation (22 transactions 2000-2025) and Qualcomm Incorporated (13 transactions 2011-2025) in the AI chip sector. The obtained results show that the methods derived from the analysis of the EDA industry transfer effectively to adjacent technology domains, with three key findings. Table 12 reports the comparison of the performances of the methods applied to the three datasets.

Metric	EDA Industry (N=661)	NVIDIA (N=22)	Qualcomm (N=13)
Algorithm	Reverse Hybrid	DBSCAN	DBSCAN
Silhouette Score	0.98	0.89	0.88
Davies-Bouldin Index	0.12	0.18	0.18
Dense Clusters Found	5	2	1
MinPts (via $\ln(N)$ rule)	10	3	3
Avg TechImpactScore	5.96	9.45	9.47
Innovation Factor	--	7.5/10	4.7/10
SSA Revenue Forecast 2030	--	\$545B	\$94B

Table 12 - Cross-Domain Validation Summary

DBSCAN parameter is adjusted based on the guideline $\text{MinPts} = \ln(N)$ and the clustering performance is examined on three different data scales from 661-transaction industry dataset to 13-transaction company dataset. The obtained clustering granularity is appropriate for each dataset and the obtained silhouette scores

are very high (between 0.88 and 0.98). The decrease in silhouette scores when moving to smaller datasets is expected given the reduced sample size rather than indicating a methodological limitation.

As a metric, TechImpactScore is meant to provide a comparable assessment of the impact of M&A across industries. The near identical average TechImpactScores of 9.45 for NVIDIA and 9.47 for Qualcomm, despite the fact that the two companies have pursued very different M&A strategies, demonstrates that the four components capture a generalizable dimension of individual acquisition quality. The Innovation Factor differentiates sharply between the two companies, with a score of 7.5 for NVIDIA versus 4.7 for Qualcomm, confirming that overall M&A strategy quality and individual transaction quality are distinct phenomena. NVIDIA's two-phase acquisition strategy (2019-2022 for AI infrastructure and 2024-2025 for AI software), provided significantly greater Innovation Factors than the single-phase focused acquisition approach of Qualcomm, even if both transaction types had relatively equal quality. Therefore, relative to raw acquisition quality, the Innovation Factor could provide a strategic evaluation of technology through capturing timing and evaluating beyond simply quality of acquisitions across the technology ecosystem.

The SSA forecasting technique developed for analysis of M&A waves has been used to evaluate revenue forecasting. NVIDIA's 2030 revenue is projected to be \$545 billion (95% CI = \$440B - \$650B) with Qualcomm's 2030 revenue expected to be \$94 billion (95% CI = \$81B - \$107B). These findings provide support for the generalizability of the SSA mathematical framework across technology industry time series categories.

SCIENTIFIC, APPLICATION-SCIENTIFIC AND APPLICATION CONTRIBUTIONS OF THE THESIS

The contributions obtained from the research work done in this doctoral thesis are grouped into three categories: scientific contributions, scientific-application contributions and application contributions.

The Scientific Contributions of this thesis are as follows:

Contribution 1 - Reverse Hybrid Clustering algorithm, a novel 3-stage clustering algorithm that first applies DBSCAN and then K-Means and reintegrates the noise into the refined clusters. This inverted order of operations preserves the natural density patterns in the data related to M&A activity, which are generally destroyed by centroid-first hybrids. Evaluation of the results on the EDA M&A dataset showed a maximum average silhouette score of 0.98. Compared to this, the silhouette score of simple K-Means was 0.72, of DBSCAN 0.85, of forward hybrid clustering 0.72 and of Gaussian Mixture Models 0.68.

Contribution 2 - Five M&A cluster archetypes. These archetypes were extracted from the application of the Reverse Hybrid Clustering algorithm to 661 EDA M&A transactions from 1975 to 2025. The results show five distinct strategic archetypes (High-Tech Targets, Growth-Stage Scale, Mature Portfolio, Young Startup Tuck-ins and Rapid High-Value), forming a new empirical taxonomy of M&A behaviour in the technology sector. The results are validated using one-way ANOVA ($F = 175.6$, $p < 0.001$), Kruskal-Wallis ($H = 46.3$, $p < 0.001$) and permutation test ($p = 0.044$). Bootstrap resampling test shows that the resulting clusters are stable with a mean ARI of 0.94 across 100 iterations.

Contribution 3 - TechImpactScore composite metric. The TechImpactScore composite metric quantifies the technological impact of an M&A transaction by combining patent score, deal value, company age and acquirer frequency, with weights (0.35, 0.30, 0.20, 0.15) derived from PCA on the 661-transaction dataset. The metric shows convergent validity through correlation with independent deal-value data (Pearson $r = 0.71$, $p < 0.001$), and ablation studies show a 34.7% drop in silhouette score when it is removed from the feature vector.

Contribution 4 - SSA-based M&A wave forecasting. SSA has been adapted to M&A transaction time series for the first time in the corporate-transaction literature, with three domain-specific design choices: a 12-quarter embedding window, three retained eigentriples, and Holt-Winters exponential smoothing for confidence-interval estimation. The decomposition recovers all four historically documented EDA M&A waves and forecasts a fifth wave peaking in 2027 to 2028.

Contribution 5 - Critical risk window for EDA company viability. A multi-dimensional Cox Proportional Hazards model has been built for EDA company viability that integrates cluster membership, TechImpactScore, network centrality, company age and technology sector. The model reaches a concordance index of 0.78 and identifies a critical acquisition-risk window between the third and seventh year of a company's life.

Contribution 6 - Network and concentration analysis of the EDA acquisition ecosystem. Graph-theoretic network analysis of the EDA acquisition ecosystem identifies three major acquisition communities centred on Synopsys, Cadence and Mentor Graphics / Siemens EDA, and quantifies a steady increase in market concentration measured by the Herfindahl-Hirschman Index, which rises from approximately 0.08 in the 1990s (unconcentrated) to 0.22 in the 2020s (moderately concentrated)..

The Application-Scientific Contributions of this thesis are as follows:

Contribution 7 - Semi-automated multi-agent M&A discovery system. A semi-automated multi-agent system has been built for M&A database expansion that combines LLM-based information extraction with human-in-the-loop validation, using four specialised agents (RSS Feed Monitor, Google Custom Search, OpenAI Extraction, Company Deduplication) and a novel three-level adaptive rejection mechanism that improves precision over time without model retraining. Operational testing shows 87 percent coverage and 77 percent precision at an operating cost below five USD per month.

Contribution 8 - Innovation Factor cross-domain metric. An Innovation Factor composite metric has been developed for cross-industry comparison of M&A strategies. Application to NVIDIA (Innovation Factor 7.5/10) and Qualcomm (4.7/10) confirms that the metric distinguishes proactive vertical-integration strategies from reactive horizontal-diversification strategies even when individual acquisition quality is comparable, validating the analytical framework beyond EDA.

The Application Contributions of this thesis are as follows:

Contribution 9 - EDA-DYV online research platform. The EDA-DYV online research platform, integrates every analytical method developed in this thesis into a single web interface with a three tier architecture (C# .NET in backend with Entity Framework, and React with D3.js visualizations in Front, SQL Server as database). The platform shows clustering analysis, wave forecasting, survival analysis, network visualisations, M&A timeline analysis and the semi automated discovery system and is deployed at <https://edadynamics.com>.

LIST OF PUBLICATIONS ON THE DISSERTATION

1. Zylfiu B., Marinova G., and Hajrizi E., "Design of an Online Platform EDA-DYVI to study the Dynamics and Forecast the Viability of EDA Companies", 14th International Scientific Conference on Development and Modernization of Manufacturing, University of Bihać, Faculty of Technical Engineering, Bihać, Bosnia and Herzegovina, 20-22 September 2023, pp. 634-640, ISSN: 2566-3283, Y. 26, No. 14, Electronic Proceedings, Conference Paper Available: <http://rim.tfb.unbi.ba/common/RIM2023v2.pdf>
2. Kostova K., Menxhiqi L., Zylfiu B., and Marinova G., "Review of Artificial Intelligence Implementation in Electronic Design Automation Methods and Tools," in Proc. UBT International Conference on Computer Science and Communication Engineering (UBT IC-CS), UBT Lipjan, Kosovo, Oct. 28-29, 2023, ISBN: 978-9951-550-95-6, DOI: 10.33107/ubt-ic.2023.284. Conference Paper Available: <https://knowledgecenter.ubt-uni.net/cgi/viewcontent.cgi?article=4527&context=conference>
3. Zylfiu B. and Marinova G., "A Web-Based Application in EDA-DYV Platform for Analyzing Mergers and Acquisitions in the EDA," IFAC-PapersOnLine, vol. 58, no. 3, pp. 193-197, 2024. 22nd IFAC Conference on Technology, Culture and International Stability (TECIS 2024), Waterford, Ireland, May 29-31, 2024, ISSN: 2405-8963, DOI: 10.1016/j.ifacol.2024.07.149. Conference Paper Indexed in Scopus, Web of Science <https://www.sciencedirect.com/science/article/pii/S2405896324002325>
<https://www.webofscience.com/wos/author/record/18514321>
4. Zylfiu B., and Marinova G., "AI-Driven M&A Forecasting: Singular Spectrum Analysis Model in the EDA Industry," 2025 60th International Scientific Conference on Information, Communication and Energy Systems and Technologies (ICEST), Ohrid, North Macedonia, 2025, pp. 1-4, ISBN: 979-8-3315-2655-9, DOI: 10.1109/ICEST66328.2025.11098239. Conference Paper Indexed in IEEE Xplore <https://ieeexplore.ieee.org/document/11098239>
5. Zylfiu B., Marinova G., Hajrizi E., and Qehaja B., "Cluster Analysis of Merger and Acquisition Patterns in the Electronic Design Automation Industry Using Machine Learning Techniques," International Journal of Innovative Technology and Interdisciplinary Sciences (IJITIS), vol. 8, no. 3, pp. 784-817, 2025, ISSN: 2613-7305, DOI: 10.1515/IJITIS.2025.8.3.784-817. Journal Paper Indexed in Scopus <https://www.scopus.com/pages/publications/105020749400>
<https://journals.tultech.eu/index.php/ijitis/article/view/336>
<https://www.researchgate.net/publication/397059352>
6. Zylfiu B., "An Intelligent Multi-Algorithm Integration Framework for Automated M&A Decision Support in Technology-Intensive Industries", in Proceedings of the International Conference on Intelligent Systems and New Applications (ICISNA'25), Vol. 3, Antalya, Turkey, Dec. 12-14, 2025, pp. 1-7, ISSN: 3023-6193, DOI: 10.58190/icisna.2025.133. Conference Paper Available: <https://proceedings.icisna.org/conf/index.php/ICISNA/article/view/112>
7. Zylfiu B. and Marinova G., "Implementing a Semi-Automated M&A Discovery Module for the EDA-DYV Research Platform," 1st International Conference on Industrial and Systems Engineering (ICISE 2026), Tirana, Albania, 16-17 April 2026, Conference Paper (Accepted, in press) Indexed in Springer
8. Zylfiu B. and Marinova G., "Comparative Analysis of AI Chip Innovation Strategies: DBSCAN Clustering and SSA Forecasting Applied to Diverse M&A Patterns," 1st International Conference on Industrial and Systems Engineering (ICISE 2026), Tirana, Albania, 16-17 April 2026, Conference Paper (Accepted, in press) Indexed in Springer

SUMMARY

METHODS AND ONLINE TOOL FOR STUDYING THE DYNAMICS AND EVALUATING THE VIABILITY OF ELECTRONIC DESIGN AUTOMATION COMPANIES IN TIMES OF TECHNOLOGICAL DISRUPTION

Over the past fifty years, mergers and acquisitions have reshaped the electronic design automation (EDA) industry. What began in 1980 as a crowded field of small, specialist tool vendors is today a market held by three firms: Synopsys, Cadence Design Systems and Siemens EDA. Most studies of how this happened stop at description. They lean on case studies, hand-sorted industry categories, and a few summary statistics, and they leave the harder questions untouched. No one has analysed EDA acquisitions with machine learning. There are no models that forecast when the next wave of deals will arrive, or that judge whether a given company is likely to survive one. Nor is there a shared online tool that would let another researcher load the data and check the results. This research starts from those gaps. The question it sets out to answer is how the dynamics of EDA mergers and acquisitions can be studied, and the viability of the companies involved assessed, with quantitative methods that stand up to statistical testing and still work outside this one study.

The work answers that question with a single framework built on solid ground: a dataset of 661 confirmed EDA transactions, checked and cleaned, covering 1975 to 2025. Six methods sit inside it, each built for a distinct purpose. The first, Reverse Hybrid Clustering, runs DBSCAN to find the dense cores and the outliers, then K-Means to sharpen those cores, and finally folds the outliers back in. Running the steps in that order keeps the density patterns that a centroid-first method would flatten. The second is the TechImpactScore, a single number that captures the technological weight of a deal from its patent, financial, temporal and acquirer features; the weights come from principal component analysis on the data rather than being set in advance. Third, Singular Spectrum Analysis breaks the transaction time series into trend, cycle and noise so that acquisition waves can be detected and projected forward. A Cox proportional hazards model handles company viability. Network analysis maps who acquires whom and how concentrated the market has become. Finally, a semi-automated system of language-model agents grows the dataset, with a researcher checking its output. All six live in one place, the EDA-DYV online platform, which was then tested beyond the EDA setting, on the adjacent AI chip sector.

The clustering comes first. Reverse Hybrid Clustering reached an average silhouette of 0.98, a wide margin over the alternatives: K-Means and the forward hybrid both sat at 0.72, DBSCAN at 0.85, and Gaussian Mixture Models at 0.68. The deals settled into five strategic archetypes, namely High-Tech Targets, Growth-Stage Scale, Mature Portfolio, Young Startup Tuck-ins and Rapid High-Value. The obvious worry with groups this clean is that they reflect the algorithm rather than the data. Four independent tests argue otherwise: one-way ANOVA returned $F = 175.6$ ($p < 0.001$), a Kruskal-Wallis test and a permutation test pointed the same way, and bootstrap resampling held the clusters steady at a mean Adjusted Rand Index of 0.94 across 100 runs. The TechImpactScore proved useful on two counts. It lined up with independent deal-value data at $r = 0.71$ ($p < 0.001$), and removing it from the feature set cost 34.7 percent of cluster quality. Singular Spectrum Analysis recovered all four merger waves the industry already recognises and pointed to a fifth, cresting around 2027 to 2028. With a concordance index of 0.78, the Cox model placed the greatest acquisition risk in the years three to seven of a company's life. The acquisition network resolved into three communities, centred on Synopsys, Cadence and Mentor Graphics / Siemens EDA, and concentration rose over time, the Herfindahl-Hirschman Index moving from roughly 0.08 in the 1990s to 0.22 in the 2020s. The approach also held up away from EDA: applied to the AI chip makers, it set NVIDIA, at an Innovation Factor of 7.5 out of 10, apart from Qualcomm at 4.7. The discovery system, last, reached 87 percent coverage at 77 percent precision, and the whole pipeline now runs live and is publicly accessible at edadynamics.com.

This thesis set out to build a coherent set of methods, together with an online platform to house them, for studying the dynamics of the EDA industry and the viability of its companies through periods of technological

disruption. That objective has been met. Each method holds up under validation, and each result can be reproduced. What remains are the questions the work opens rather than closes. The first concerns what sits behind the waves: future study could go past detecting acquisition waves and begin explaining them, separating the surges set off by genuine technological disruption from those that trace back to consolidation or to favourable financing. A second is whether the TechImpactScore keeps its value after the deal closes, judged against what an acquirer genuinely gains from an acquisition in the years that follow. The methods themselves could reach further as well, into biotechnology, financial technology and other sectors, while real-time data feeds and deep-learning feature extraction would broaden what they cover and lessen their reliance on manual oversight. These questions are taken up in mor