Artificial Intelligence Review (2024) 57:44
https://doi.org/10.1007/s10462-023-10664-4

Modeling and prediction of business success: a survey

Divya Gangwani[1] · Xingquan Zhu[1]

Published online: 9 February 2024

Abstract

Businesses are the driving force behind economic systems and are the lifeblood of the community. A business shares striking similarity to a living organism, including birth, infancy, rising, prosperity, and falling. The success of a business is not only important to the owners, but is also critical to the regional/domestic economic system, or even the global economy. Recent years have witnessed many new emerging businesses with tremendous success, such as Google, Apple, Facebook etc., yet millions of businesses also fail or fade out within a rather short period of time. Finding patterns/factors connected to the business rise and fall remains a long lasting question puzzling many economists, entrepreneurs, and government officials. Recent advancement in artificial intelligence, especially machine learning, has lend researchers powers to use data to model and predict business success. However, due to data driven nature of all machine learning methods, existing approaches are rather domain-driven and ad-hoc in their design and validations. In this paper, we propose a systematic review of modeling and prediction of business success. We first outline a triangle framework to showcase three parities connected to the business: Investment-Business-Market (IBM). After that, we align features into three main categories, each of which is focused on modeling a business from a particular perspective, such as sales, management, innovation etc., and further summarize different types of machine learning and deep learning methods for business modeling and prediction. The survey provides a comprehensive review of computational approaches for business performance modeling and prediction.

Keywords Business success · Features · Supervised learning · Clustering · Machine learning

Divya Gangwani and Xingquan Zhu contributed equally to this work.

✉ Xingquan Zhu
xzhu3@fau.edu

Divya Gangwani
dgangwani2017@fau.edu

1 Department of Electrical Engineering and Computer Science, Florida Atlantic University, Boca Raton, FL 33431, USA

1 Introduction

In today's global landscape, a surge in entrepreneurship is evident, as new businesses and entrepreneurs emerge worldwide, driven by the embrace of entrepreneurial spirit and the exploration of diverse business opportunities (Zane 2023). The phenomenon is triggered by many factors, including government support, technical innovations, cultural and economic factors encouraging people to follow their passion. Every entrepreneur and private investor aspires to build a unicorn like Uber and Facebook, which have transformed the way we think about businesses and have a tremendous impact on society. Nevertheless, each business has its own risks of failure along with the probability of success. Potential risks such as not fulfilling the financial goals, poor management strategies, wrong hiring, and marketing mishaps are the most common reasons for business failures (Jaafari 2001). With the rise of startups and their impact on the economy, entrepreneurs, investors, and decision markers are in need of effective methods to analyze business data from different perspectives.

Finding relevant factors affecting business fluctuations has been a challenging task due to constant evolving of the technologies, competitive market, and industrial revolutions. Recent studies have focused on factors such as probability of firms' merger and acquisitions, monetary factors that lead to business success and investments in the companies to have IPO status (Ross et al. 2021; Breitzman and Thomas 2002). Nevertheless, these studies have essentially limitations, because they focus on a particular method or a limited number of factors.

In order to make valid prediction of business success, it is essential to consider relevant theories and factors contributing to the business rise and fall. Every business follows a systematic routine involving various stages or series of events known as the business life cycle, as illustrated in Fig. 3. Similar to the life cycle of living organisms, businesses strive to achieve success through different stages. An abundant number of theories (Mankiw 1995) support the importance of the business cycle in exploring major factors minimizing the effects of fluctuations and attaining success (Fig. 1).

Motivated by existing business studies and computer science research, this paper aims to leverage theories of business fluctuations to study essential features and factors relevant to business success, and showcase a systematic framework that categorizes relevant factors into three main entities for the prediction of business success using machine learning and deep learning methods. We define business success from the computational point of view such that the approach will provide a verifiable solution for predicting business success.

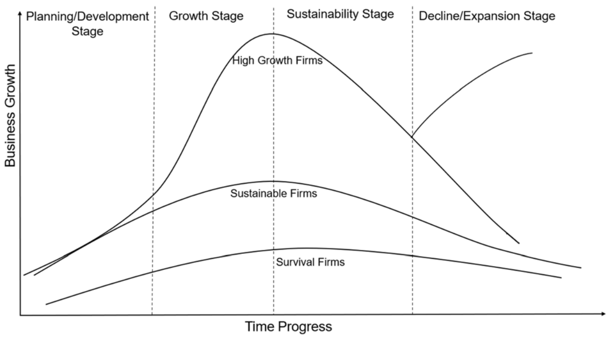

Fig. 1 A conceptual view of five main stages of a business life cycle

Our approach will benefit young entrepreneurs, investors, and even unicorn companies who are constantly seeking methods to predict business success.

1.1 Business success

Indeed, startups and established firms have unique information about their finances, investments, funding information, market growth, innovations, employee details etc., which hold keys determining their long trajectory and success. Although business success can be assessed from different perspectives, from computational point of view, the success is based on the survival and the growth of the company, its capability to withstand ups and downs in the market without having to exit. Two key milestones exist to measure the business success: (1) A company going public i.e. probability of getting an IPO. (2) A company being acquired by a larger company or being merged with another company of the same level known as M &A.

1.2 Business success prediction

Modeling and predicting business success has been an interesting and challenging topic for researchers with many solutions focused on using statistical methods. Over the last decade, machine learning methods (Ünal 2019; Yu et al. 2018; Gangwani et al. 2020) have proven to be reliable and outperformed traditional methods business success prediction. Many supervised machine learning methods have been used in previous studies. For example, random forest, Naive Bayes, Support Vector Machine, Decision Trees and Logistic Regression were used based on market, financial, and demographic features of the public and private firms (Ravisankar et al. 2011; Wei et al. 2008; Saura et al. 2021). In addition, unsupervised algorithms such as k-means clustering have been widely used due to its advantage of generalizing to new examples well and scaling to large datasets (Kakati 2003; Heidari et al. 2021).

In addition to classical machine learning approaches, deep learning methods have also been used to enhance feature extraction (Borchert et al. 2022), e.g. extracting textual data from social media sites for prediction. Financial sectors, such as Banks and Stock Market industry frequently use deep learning methods to analyze customer sentiments or financial reviews in business modeling.

Indeed, many methods have been proposed for business success prediction, and several surveys (Lin et al. 2011; Devi and Radhika 2018; Qu et al. 2019) have summarized machine learning models and highlighted relevant features, but existing approaches are often domain-driven and ad-hoc in their design and validations. Due to the abundance of available data, obtaining insights into a business and accurately predicting its performance can be challenging, especially when considering complicated factors involved in the business operating, e.g. products, human resources, market, investment, management etc.

In this paper, we review important features and create a framework for modeling and predicting business success. A unique component of our work is the investment-business-market framework which characterizes three major key components related to the business. Based on this framework, our survey further outlines various features and models related to the business success prediction. The survey is beneficial for stakeholders, researchers, entrepreneurs, and investors for them to identify determinants of success in various business contexts.

1.3 Our contribution

Our study aims to provide a comprehensive up-to-date literature review on business success prediction. Existing studies are focused on limited factors such as organization details or investments made in specific industry or sector. Our study helps close the gap by providing a systematic framework to thoroughly review case studies, articles, and theories related to business fluctuations, factors, and variables contributing to a company's success or failure. A review of learning methods, data, and performance metrics further outlines the whole eco-system of using machine learning for business success prediction. The main contribution of this study is listed as follows:

- We conduct a comprehensive selection process for modeling and prediction of business success. Our study includes various articles and research based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) method, which identifies important records pertaining to our study and excludes records based on the exclusion rules. A screening process is carried out to collect relevant articles and remove duplicate or incomplete ones to improve the quality of the study. The selection criteria and the selection process are reported in Fig. 2.
- An Investment-Business-Market (IBM) triangle framework is used to summarize critical factors responsible for analyzing business success. This framework serves as a general skeleton covering vital features related to business life cycle and operation.
- We provide a detailed study of major features used for predicting business success, categorize these features according to the three main parities of the IBM triangle, and explain how these features are extracted and modeled based on different business angles.
- An extensive survey of machine learning models for business success prediction provides a landscape for researchers, investors, and entrepreneurs to understand the state-of-the-art. It will also allow them to pivot in a timely manner so that financial resources are utilized wisely.

The rest of this paper is organized as follows. Section 2 introduces theories of business cycle and IBM triangle framework to study factors and criteria for business success. Section 3 categorizes features for business modeling based on the IBM triangle. A review of machine learning and deep learning models for business success prediction is reported in Sect. 4. Section 5 lists public datasets and performance metrics for modeling business prediction, and Sect. 6 concludes the paper.

2 Business cycle theories and IBM triangle

2.1 Theories of business cycle

For decades, business evolution, along with industry revolutions and bankruptcies, has given economists and researchers tremendous opportunities to observe driven factors, and draw important business cycle theories from different perspectives. A business cycle refers to the periodic ups and downs that an economy faces during different stages

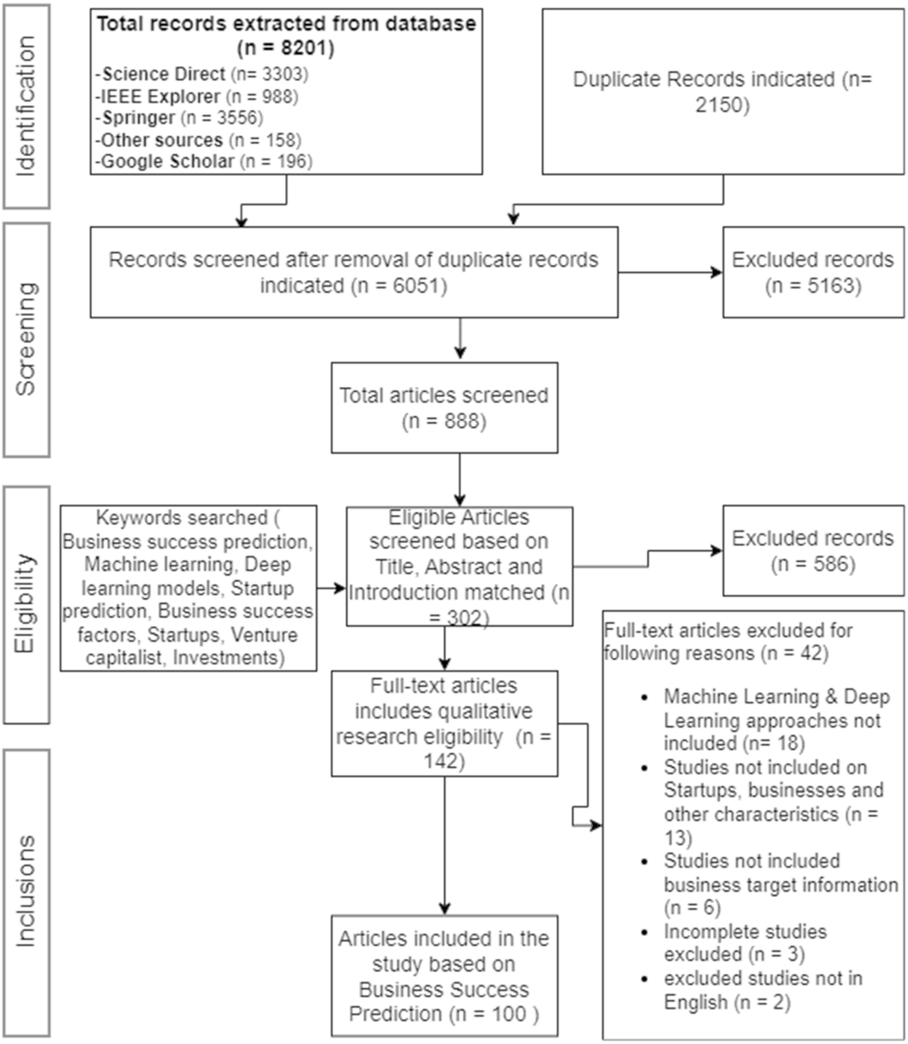

Fig. 2 A study based on selection process of reviewed articles

of business leading to the decline or expansion stage. These fluctuations or changes in the cycle can occur over a period of time that can cause a sudden rise in the product pricing or decline in profit. Several well established theories, such as Keynesian's theory (Mankiw 1995) related to business cycle, have emerged and greatly influence how modern business work.

The foundation behind creating the business cycle relies on economists who periodically studied the real-time GDP and investments over the years to develop a theory behind defining the four main stages of the business cycle. In the early 20th century, it was observed that businesses had seen a tremendous rise in profits and growth of the business for several years due to the technological changes in the economy, which was followed by a drastic decline that came as a shock to many businesses. The shocks such as innovations, investments, uncertainty, over-production of goods, and environmental factors affected the rise and fall of the business (Cerra et al. 2020).

Many researchers propose different business theories to analyze main factors leading businesses toward success or failure. These factors can be measurable or non-measurable such as environmental factors. Focusing on measurable factors helps economists or policymakers make predictions based on innovations, investments, profits, and sales to save businesses from the consequences of cycle fluctuations. Table 1 summarizes eight business theories and factors related to cycle fluctuations. For example, the theory of innovations (Schumpeter 2017) in business suggests that introducing new technology and new techniques in selling a product influences continuous growth in the company and leads to long-term returns on investment. Continuous innovations increase the risk of uncertainties; however, these innovations help the business compete with its competition, therefore assuring a uniform rate of profit over time. A focus on the theory of human capital in organizations (Brüderl et al. 1992) demonstrates competitive factors that influence the success of new organizations. Business demographics and founders' skills provide an impact on how an organization would survive during the shifts in the business cycle. A newly found organization has very little competition in the market and hence can utilize the human capital theory to capture the market's attention. A major factor that draws the attention of entrepreneurs is the market for the business. Many theories highlight marketing to be an important aspect to capture investors as well as create customer product engagement for long-term growth (Kleinaltenkamp and Jacob 2002; Hunt 2013). Allocating new marketing strategies and resources not only brings new investors into the business but also attracts customers via Business-to Business (B2B) or Business-to-Customer (B2C) relations thereby providing an advantage for the organizations to stay on top of the businesses during economical shifts. Fluctuations in investment are critical to the theory of the business cycle and must be considered when focusing on business success. The impact of investments and finances provides a disturbance in the business cycle trends and is the main reason for shocks. For example, if the investments increase, the productivity also increases with a rise in business capital, thereby increasing the profit and returns on investments. On the other hand, if the investments decrease, the business output is greatly affected hence reducing the profit as well as reducing the employment opportunity for business (Greenwood and Jovanovic 1990; Justiniano et al. 2010) Another study, on the role of strategic management in the business cycle, explains that strategic planning forces the organization to think differently in improving its capital expenditure, supply chain, human resources, business operations, and product pricing. It unfolds two major aspects which help to answer the questions 1) When should the organization make changes and 2) how to implement those changes. When applied within the organization's workforce in a timely manner throughout the stages of the business cycle, strategic planning can minimize the effect of fluctuations.

Combining theories related to major factors influencing business cycle provides a roadmap on building solutions to help businesses succeed or grow in competitive business environments. Based on business theories shown in Fig. 3, we propose a framework to group all factors into three main categories: investments, business, and market, which led to the creation of the IBM triangle in Fig. 4.

2.1.1 Case studies on fortune 1000 companies

The practical values of the business cycle theories are further evident by the Fortune 1000 companies' success. For example, in 2022, Walmart was ranked No. 1 in the Fortune

Table 1 Summary of business cycle theories. The table lists eight theories summarizing factors/hypothesis about business evolution

Theory Name	Factors	Hypothesis
Innovation Theory of Profit (ITP) Schumpeter (2017)	Business profit, Growth, Innovation	Creating innovation was the first step on the path to success and economic profits for entrepreneurs
Business Theory of Human Capital (BTHC) Brüderl et al. (1992)	Founder demographics, Organizational survival, Human resource, Business Capital, Product development	Business theory influence success or survival of newly found organizations focusing on business demographics, founders details and market capturing
Theory of Business-to-business Marketing (B2BM) Kleinaltenkamp and Jacob (2002)	Market strategy, Product development, Competitive advantage over b2b, Investor growth	Theory of b2b marketing focuses on attracting new customers in buyers and seller's market thereby capturing new investors into business
Resource-Advantage theory of Marketing (R-AM) Hunt (2013)	Competitive advantage to firms, Firms success with marketing strategies, B2b and B2c competition	Resource allocation theory provides a competitive advantage over business marketing of successful firms
Strategic Marketing Theory for Business Performance (SMBP) Morgan (2012)	Acquiring resources, Investments opportunity, Market capability, Business operations	Strategic market in terms of investments, resources and organizational capabilities plays an important role in the growth or success of business
Theory of Link between Finance and Corporate Growth (LFCG) Greenwood and Jovanovic (1990)	Financial investments, Economic growth, Net income, Profitability	The evidence from cross country studies suggest a strong link between finance and corporate growth
Investment Shocks and Business Cycle (ISBC) Justiniano et al. (2010)	Stock price inflation, Investment output, Marginal wages	Investment shocks are the main reason for fluctuations in the business cycle
Multiple-Accelerator Interaction Model (MAIM) Samuelson (1939)	Private Investments, Aggregated income, Economic distribution of wealth, Income consumption	Investment multiplier can affect the consumption of wealth as well as accelerate the distribution of wealth across the economy

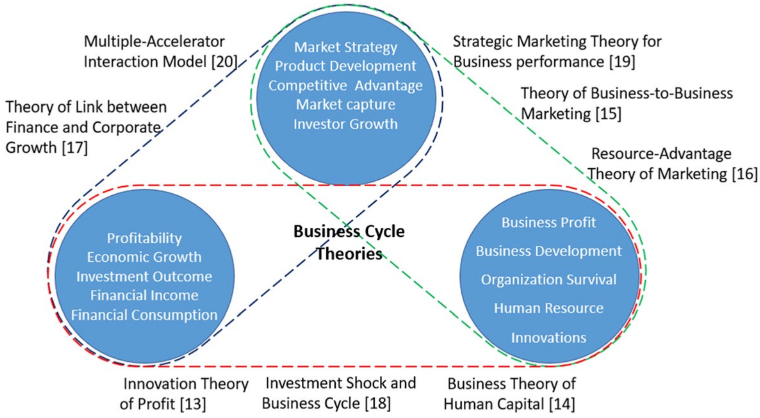

Fig. 3 Theories of Business Cycle and relevant factors behind the theories

1000 companies within the U.S. followed by Amazon, Apple, CVS etc. KENTON (2022). Walmart has around 2.3 million employees working in the company with total revenue of $572,754 million generated. Based on statistics, it is observed that the rise in hiring of employees, investing in the right products and selecting demographic locations close to the product sourcing are top factors responsible for generating higher growth revenue.

A survey conducted by CEOs of large companies such as Tiffany and Chick-fil-A, studies the importance of measuring innovations and developing metrics effectively measuring innovation in the ever-changing market. Monthly evaluation of KPIs with financial indicators was used to measure how new ideas and innovations were set in place (Tucker 2021). These ideas and innovations start with hiring the right CEO for the company. According to a Harvard Business Review (HBR) article in 2017 (Editors 2017), top-100 leaders were ranked as world's best performing CEO's of their company. Metrics such as financial profits and non-financial indicators were used in evaluating and delivering results. Company's growth in terms of innovations, market capture, market expansion, proximity to the product sourcing, product delivery and Returns on Investments (ROI) was measured. CEO's such as Pablo Isla, Jeff Bezos etc. were among the top ones capable of keeping their businesses at the top.

The above evidence suggests that creating a new vision with innovative ideas and generating a new business model is the key to success. Regardless of the market sector, large companies are actively using machine learning and AI to provide next-level products and services to customers. Alibaba (Marr 2019) is one the leading e-commerce company utilizing Natural Language Processing (NLP) to generate product descriptions and utilizing forecasting models to predict customer-product engagement. Alphabet, a parent company of Google, relies heavily on deep learning algorithms to promote self-driving cars. Tech giants like Amazon, Microsoft, IBM, Tencent (a Chinese social media company) use machine learning, AI and cloud platforms to promote customer satisfaction, enhancing employee capabilities, product distribution, understanding customer engagement, product innovations and so on.

Considering above mentioned factors, there exists a correlation between features such as market, business demographics, product, investments and innovations useful for predictive models. Depending on the type of business such as large, small, or medium firms the

Table 2 Summary of important factors to the business growth

Features/businesses	Fortune 1000	Medium firms	Small firms	Startups
Business innovation	✓	✓	✓	✓
Human resource	✓	✓	✓	✓
Demographics	✓	✓	✓	✓
Investments	✓	✓	✓	✓
ROI	✓	–	–	–
Rate of market scope	✓	✓	–	✓
Product innovation	✓	✓	✓	✓
Financial capability	✓	✓	–	–
Market growth	✓	–	–	–
VC Funds	✓	✓	✓	–
Technology	✓	✓	–	✓
Profit	✓	–	–	–

availability of the features/factors remains inconsistent as summarized in Table 2. Even with this knowledge, it is still challenging to quantify and measure each component and analyze the criteria for business success.

2.2 Investment-business-market relationship for business modeling

Existing studies on business success prediction primarily focus on certain aspects of features but fail to consider theories and factors leading to fluctuations in the business economy (Weill 1992; Mason and Harrison 2002; Lee et al. 2019; Santisteban and Mauricio 2017; Wan 2010). Motivated by the business cycle theories and relevant factors outlined in Fig. 3 and our previous study in exploring investor-business-market interplay for business success prediction (Gangwani et al. 2023), an Investment-Business-Market triangle framework, as shown in Fig. 4, is used to summarize three aspects critical to the business success.

The IBM triangle is intended to serve as an umbrella framework for us to review various factors and theories related to business fluctuations.

Meanwhile, it also helps answer important questions about predictive models, such as, what are the key features in predicting business success, how do these features

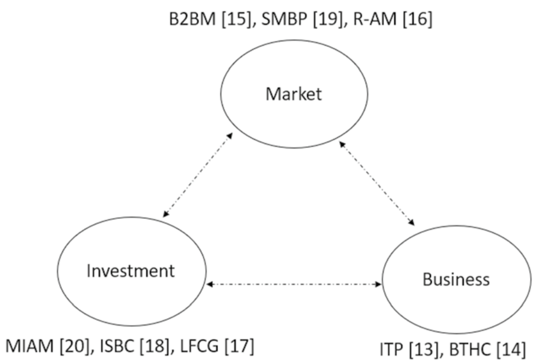

Fig. 4 Investment-business-market triangle framework summarizing Investment, Business, and Market triangular relationship (Gangwani et al. 2023)

interrelate with each other, and most importantly, what methods are available to predict business success?

Depending on the data availability, IBM triangle provides a road map for feature selection and feature engineering required during model building. There is a growing amount of literature highlighting the relationship between investment, business and market. The literature can be divided into three strands to show the relationship between these entities:

- *Investment and Business relationship*: Investments are assets or funds invested in a company with an expectation of long-term growth specifically through M &A or becoming an IPO (Pan et al. 2018). Ideally, the investment journey starts with planning and strategy. There is always a time frame associated with any investments which can be fruitful in long-term planning. Investments may be financial, such as foreign investment, or they may involve innovation or a company's obtaining patents. It may also be based on the business's potential or economic growth. Rai et al. (1997) show the relationship between technological investments and business performance. Wan (2010) conducted a survey to demonstrate the relationship between foreign investment and the economic growth of the company. Many researchers show that investments in patents can have high business potential as the investors can hold or maintain their rights over the patents and gain profit (Ernst 2001; Choi et al. 2020; Yuan et al. 2020). These types of investment and business relationships show that there are high chances of success when a business finds the right paths to move forward with a plan to invest and gain maximum profit.
- *Market and Business relationship*: A healthy market consists of buyers and sellers to exchange products or services. Therefore, the market has a tremendous impact on business performance. Market orientation is a key factor in defining where a business reacts to the demand in the market. Research has shown (Greenley 1995; Iyer et al. 2019; Al-Henzab et al. 2018; Sandvik and Sandvik 2003; Brik et al. 2011) that market orientation factors, such as market shifts, technological changes, product innovation and brand management, social responsibility, have a positive impact on business performance. Market resources such as *(price, advertisement, distribution)* and market knowledge capabilities (Morgan 2012; Hou and Chien 2010) also play a significant role in business performance.
- *Investment and Market relationship*: Investment is directly related to the market. Market increasing, such as a stock market, often promotes product sales *etc.,* increased investment. Investors tend to invest in stocks showing clear growth in the company's stocks and shares (Vui et al. 2013; Barro 1990). Based on this analysis, we can observe that investments shift to follow the market. Dot-com bubble, blockchain, and cryptocurrency exemplify investment tendency on the market (Gangwani et al. 2021; Hawkins 2004; Crain 2014; Bouri et al. 2019; Akyildirim et al. 2021). By observing the trends and the studies, we can find that investment is always biased towards different markets.

The above multifaceted correlation shows interrelation between seemly complicated factors behind the business operation. The IBM triangle entities are considered to provide a basis for modeling and predicting business success using machine learning algorithms. Table 3 summarizes business, market, and investment-related features and sub-features and their strength *vs.* weakness for predictive modeling.

Table 3 Summary of features related to business success

Feature	Sub-feature	Strength	Weakness
Business Features	Demographics, administrations,HR, education, product details, social media publicity,finances	A clear example of business features used for evaluation	Difficult to obtain complete information
Market Features	Market growth,sector, customer satisfaction, product creation, product value	Demonstrates good KPI's. Provides temporal variance	Sectors may change with time
Investments Features	Funding amount, innovations, ROI, Patents,Hiring & training cost	Most important & easily available metrics	Huge amount of redundant data

2.3 Business success criteria and definition

Many criteria exist to evaluate critical success in the business. We select measurable criteria based on the previous studies and summarize several business success criteria and factors responsible for evaluating business success. Criteria such as project timeline, stakeholder's satisfaction etc. are not considered as measurable as it does not generate a clear outcome. Table 4 lists important criterias for business success. Financial factors are important for analyzing the business success (Altman 1968; Ünal and Ceasu 2019). Investments and funding in the company, stock market evaluation and bankruptcy are always evaluated using financial criteria of the business. Marketing characteristics (Kozielski 2019; Dibb 1998) on the other hand such as market orientation, market segmentation, etc. target a specific market based on the products in small and medium size firms to provide useful factors for forecasting the business growth.

Many researchers in the past have used managerial and entrepreneurial factors for measuring success criteria in business (Bento 2018; Pasayat et al. 2020; Kakati 2003; Gorgievski et al. 2011). These criteria have been proven to maximize growth and profit in small and mid-size companies. Product characteristics have also been important factors in analyzing the company's growth. It is essential to have a good quality product that meets customer's needs and expectations (Kakati 2003).

Considering important factors mentioned above and key components required when measuring business success, in this paper, we define business success as a capability of a company to become an IPO or be acquired or merged with another company (M &A) and receive more funds from investors or VCs. On the other hand, failure is defined as a business being formally closed or bankrupted. With this definition, machine learning tasks are to identify and distinguish companies between failure and success, i.e. a binary classification or multi-class classification task.

3 Features for business modeling

Analyzing features for business prediction is a necessary step in order to build a prediction model. Figure 5 outlines primary business model features categorized under each section of our IBM framework, as well as outlines interconnection between each sub features. Each feature category lists important sub-features useful for entrepreneurs to create a business strategy plan for the company.

3.1 Investment features

Investment features identify factors related to investments made to the business. For example, the famous Dot Com investment in *2000* was a risky move in the history of web commercialization (Crain 2014) which failed and crashed the market. Therefore, investors and stakeholders nowadays evaluate all aspects of financial features of the company to stay on top of the decision-making process. Another aspect of investment feature includes human resource and technological features. These features include essential information about the companys' monetary information which is critical for evaluating whether the company is in profit or loss. Hence, in the Table 5, we summarize

Table 4 A summary of business success factors and performance indicators

Business success factors	Performance indicators	Definition	References
Financial criteria	Investments	Money invested in the company	Song et al. (2018)
	ROI	Return on Investment	Żbikowski and Antosiuk (2021)
	Funding	Received funding from VC	Stuart and Abetti (1987)
	IPO	Stock market value of a public company	Fuertes-Callén et al. (2020)
	Seed Funding	Initial funding received	Fuertes-Callén et al. (2020)
Market characteristics	Market Growth	Potential increase in the sale of goods and services in a company	Żbikowski and Antosiuk (2021)
	Market Scope	The number of products and the variety of products needed based on their geographical location	Chandler and Hanks (1994)
	Market Need	The need of the product or services based on the location	Chandler and Hanks (1994)
	Market Competition	Companies selling similar products at similar price	Chawla et al. (1997)
Venture capitalist factors	Merger and Acquisition	Companies merging or acquiring another company for long-term growth	Chandler and Hanks (1994)
	Business Strategy and Plan	The number of acquisitions and strategic alliances with other company	Santisteban et al. (2021)
	Financial Viability	The extent to which the company can grow or potential measure of the company's finances	MacMillan et al. (1987)
	Capital Assets	The financial assets of the company	Żbikowski and Antosiuk (2021)
	Long-term Sustainability	Sustainability of the company over the years	Ernst (2001)
	Business Partners	The number of partners the company has	Santisteban et al. (2021)
	Business Valuation	Company's potential to achieve $1 billion (unicorn) valuation	Lee (2013)
Managerial factors	Business Plan	Managers plan to execute the project within the group of employees or a team	Santisteban et al. (2021)
	Team Size	The size of the team	Fuertes-Callén et al. (2020)
	Technical Experience	Technical experience of the manager accessing a team	Santisteban et al. (2021)
Product characteristics	Product Quality	How good the product is compared to the similar product in the market	Ernst (2001)
	Product Quantity	Ability to supply the huge amount of the product	Ernst (2001)
	Uniqueness of the Product	How unique the product is	Ernst (2001)

Table 4 (continued)

Business success factors	Performance indicators	Definition	References
Entrepreneur criteria	Age of Entrepreneur	The age of the founder	Kim et al. (2018)
	Skills	The skills needed for the growth of the company	Żbikowski and Antosiuk (2021)
	Desire for Success	The extent to which the founder is available to take risks	Tomy and Pardede (2018)
	Growth Ability	The ability of the founder to invest or get more funding	MacMillan et al. (1987)
	Market Experience	Experience of the founders or entrepreneur in the market	Chawla et al. (1997)

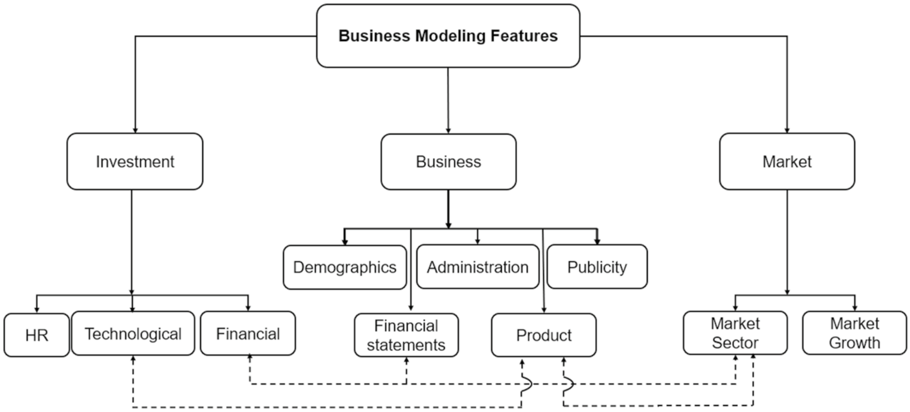

Fig. 5 A summary of main business modeling features associated to the IBM triangle. The dashed lines showing related feature subcategories

investment features into three main subcategories which is essential for describing business in terms of funding, investments and innovation made into the company.

3.1.1 Human resource (HR) features

Human Resource (HR) is one of the most valuable asset for small or large enterprises. Investments made in HR is essential to ensure the prosperity of business and changes in the market environment. Strategic investment in HR can bring a bright future to a company in terms of growth and a competitive market. Nowadays, much attention is given to the company's finances from the stakeholder's or entrepreneur's point of view to keep track of profit, loss, budget, and payroll of employees. Hence HR investments are needed in order to manage the company investments and budgeting to maintain steady growth in the market (Why now is the perfect time to invest in HR 2020).

HR investments include human capital investments such as sales made by employees, cost of hiring staff, training and educating the managers, employee and team leaders, success planning, leadership development, improving work conditions of people and providing financial and health benefits to the employees. Founder-related HR features, such as job skills, titles, and roles, are also investments made by the company. Founders are the main reason companies can reach new heights by implementing new ideas and innovations in order to bring higher returns on investments (Drábek et al. 2017; Four ways HR xxx).

3.1.2 Financial features

Financial features highlight the funding and the numeric goals of the company during the planning phase, which helps in effectively managing and running the company. Decision-makers need to assess the financial aspects of the company to identify risks or challenges associated with whether to invest in the company or not. Perboli and Arabnezhad (2021) introduced a machine learning-based decision support system that predicts mid and long-term company crises for those at risk of being declared bankrupt. Financial statements of *160,000* Italian enterprises were used as an important feature to predict companies with high chances of getting bankrupt. Apart from financial statements, monetary funds and

Table 5 A summary of Investment features and subfeatures with respect to level of difficulty (DL) to obtain, and its business relevance (BR) from low (+), medium (++), to high (+++)

Features	Sub features	Sub category of features	DL	BR
Human	HR	Human capital investment	++	+++
		Job skills	+	+++
Resource	Investments	Sales per employee	++	+
Features		Business skills competency	+++	+
		Size of capital invested	+++	++
		Cost of capital	+	+++
		Educational investment in manager	+++	++
		Cost of staff	+	+
		Managerial training investment	+++	+
	Founder information	Founders job skills	+++	+++
		Job title of founders	+	+++
		Total # of founders	+	+++
Financial	Monetary funds	Number of funds available	+	+++
		Seed funding	+	+++
		Rounds of funding	+	+++
Features		Funding goal	++	++
		Total funding	+	+++
		Amount of funding raised	+	+++
	Investors	# of investors	+	+++
		Types of investors	+	+++
		Raised amount by investors	+	+++
	Investor outcomes	Project goal	++	+
		Project category	+	++
		Project launch date	+++	+
		Return on investment	+	+++
		Merger and acquisition, IPO’s	+	+++
		Number of venture capitalist investors	+	+++
Technological	Patent	# of Patents	+	+++
		Patent classification	+++	+
Features	Innovations	Company trademark	+++	+
		Design	+++	+
		Intellectual property (IP)	+++	++
		Innovation ideas	+++	++
		Innovation target	+++	++
		Technological classification	+++	+
		Innovative participants	+++	+
		Innovation results	+++	++

financial outcomes are also responsible for predicting the success and growth of firms depending on the type and business sector.

Monetary Funds and Financial Outcomes: Monetary funds include financial investments made by investors or stakeholders in small or mid-size companies with an expectation of profit and growth in the business or financial or liquid assets set aside for investment purposes. Investors such as venture capitalists (VCs) invest in different types of firms, including startups, where they can foresee maximum ROI by either merger and acquisition (M &A) or becoming an IPO.

VCs are professional investors that invest funds into new startups or mid-size companies with an expectation to gain maximum growth by collecting financial information such as capital gain, innovation strategy, or investments made over time.

Startup firms mainly look for such investors to gain profit and sustain in the market for a longer term to eventually exit and become public. Kerr et al. (2010) use a regression discontinuity approach to analyze the role of funding in startup companies to assess the high growth potential or survival.

Because startups are relatively new to the market, evaluating them by measuring their returns on investments or other financial outcomes is difficult. Hence, other quantifiable features such as net revenue, sales, and monthly goals, known as an initial success by the investors, are used to evaluate startup growth. Apart from VCs, stakeholders also look for more information about funding, such as the amount invested by the investors, total rounds of funding, founders' information etc.

Based on the above information, companies backed by VCs are more likely to be acquired and less likely to fail in the market. This information is measured by the financial outcomes, such as the company's expected ROI. Higher ROI gives better chances for the company to be acquired or become public (IPO) and easily exit. IPOs have the maximum ROI compared to other companies, such as startups or mid-size firms.

3.1.3 Technological features

Technological features provide an upper hand in the economic growth of the business as it facilitates the venture capitalist funding required for the business to grow. Patents, innovations, or other proprietary technologies allow VC investors to control their commercial usage and provide the right to the investors to retain their benefits to the company (Innovation and Intellectual Property xxx). There are several papers which use investment information such as financial features to evaluate business performance, but technological features or strategies are not used widely because they are less quantifiable to measure. On the other hand, it is also difficult to measure technological status of a company.

Patents: Patents grant exclusive rights to investors on their innovative ideas and prevent others from using or importing the same idea for their benefit. An innovative patent brings a new horizon to the market and business planning by creating new technological development or breakthrough of a new product that has not been invented before (Ashton and Sen 1988). Therefore, we can use patents as features in business process planning as they provide unique technology information reflecting product and market development areas. Another advantage of patents is that they are very well categorized and maintained from different offices, such as the United States Patent and Trademark Office (USPTO), which maintains patent information since 1802.

Advanced technology stands better chances in developing new products in the competing market. Therefore, patents help partner selection in large firms because they are essential for research and development collaboration (R &D) (Solesvik and Gulbrandsen 2013). R &D employees bring new innovative projects and technological growth, which may result in a profitable return for small and mid-size enterprises (SMEs). On the other hand, the expenditure of R &D is huge for SMEs and their resources are limited to developing new products. Therefore, they tend to collaborate with large firms to obtain the right skills and necessary resources in developing new products (Lee et al. 2016).

Patent analysis is another form of technique that uses patent data to measure technological capabilities of the business. Such approaches are frequently used across broad area of application domains such as stock market estimation (Deng et al. 1999), merge and acquisition (M &A) (Ali-Yrkkö et al. 2005; Breitzman et al. 2002). and R &D collaborations (Teichert and Ernst 1999). An ensemble learning (Wei et al. 2008) based method summarizes five features under patent category: technological quantity, technological quality, diversity, and compatibility, to build a prediction model for M &A analysis.

Patent analysis can also assist in making technical decisions for both inventors and the R &D employees of the company. It includes information such as technical and commercial knowledge about innovations in creating new products in the company and the market strategies used in competing with similar products in the market. For successful business, it is important to have a technological road-map analyzing marketing strategies, planning and acquiring technological assets to fulfill the market needs of the future aspects of the business (Lee et al. 2009). Hence technological road map and patent analysis together play an important role in the future of the business.

Patent citation denotes the number of times a pattern has been cited by others (in other patents or publications). In many cases, a patent cites other patents with similar or related technology (Chang et al. 2009), and a patent may also be cited in scientific publications. The main idea behind patent citation is that higher citations potentially bring more significant attention (and innovations) to society, generating more competitive ideas in the market (Breitzman and Thomas 2002). Companies with more patent citations will likely possess better technological advantages and a competitive societal market. As a result, patent citation is an indicator to evaluate a firm's invention and technological creativity in the market.

Innovations: Business model innovation (BMI) (Acciarini et al. 2023) plays essential roles for firms to maintain competitive. Innovations bring new ideas or creation to the table by investors or entrepreneurs. Innovations and patents go hand in hand, as the new innovation helps the authority grant patent to investors. Although a patent must have innovative idea, in this paper, we categorize innovations as a separate measure because it can survive independently in the market without necessarily having a patent. In this subsection, we briefly explain how innovations affect the business or society.

Innovations impact the organizational culture and influence the product development in the market, hence it must be considered crucial during the business product development phase (Aksoy 2017). Organizational culture such as work ethics, social norms and role of employee influences the innovations within the group of people. Employee and manager relationship and coordination in planning, generating new ideas and creations affect the attitude toward innovations in the business. It is known that employee-manager relationships or conflicts can positively or negatively impact business (Feldman 1988). Leadership is another factor affecting the organizational innovations. Three key aspects: idea generation, risk-taking, and decision making process within the leaders play an important role in economic growth and development of a business.

- Idea generation is the process where organization takes a step forward in listening to the new ideas from employees working in the company and carefully evaluates without any criticism from others and decides whether to move forward in the direction of innovation.

- Risk-taking is an experimental stage where organizations support ideas of the employees with the possibility of achieving success in the business. Risk-taking is a step that moderates the leadership and organizational innovations in the company.
- Decision-making is the final step where leaders in an organization encourage employees to create more ideas towards innovations. Such creativity and innovations bring more success and generate new opportunities and growth in the business, regardless of whether the business is developing a product or providing new services to customers (Mokhber et al. 2018).

Innovations also affect the market and products of the business. It brings the possibility to develop new product and increases market opportunity, therefore market and product innovations go hand in hand. If leaders or employees have good marketing skills, there is also a high possibility of a new product development (Aksoy 2017). Hence, investment features directly contribute towards growth and success of the business.

3.2 Business features

Financial factors are primary focus of existing studies (Song et al. 2018; Abdullah 2021; Yazdipour and Constand 2010) but very few (Arroyo et al. 2019) emphasize on non-financial factors to predict business success. Both financial and non-financial features are responsible for evaluating successful *vs.* unsuccessful companies. Table 6 outlines business features and its subcategories in detail.

3.2.1 Administrative features

Administrative features include information about employees' work experience, salary, education, and technological skills used in their day-to-day work. It also includes supervisory information like number of investors to date, number of managers working on different projects and the team size handled by managers. All this information gives a clear idea about the company's work culture, whether the company is up-to-date with the latest technologies and provides us with the information about the company's financial capability to withstand the pressure from the outside environment. It helps entrepreneurs and investors handle business fluctuations and prepare for external factors affecting business growth.

3.2.2 Demographic features

The success of a business is influenced by demographic factors, such as the location of the company's headquarters and the products or services it offers in the market (Chin 2020). The proximity of a company's headquarters to its market sector is a leading factor associated with company growth and profitability (Boasson and MacPherson 2001). Moreover, success in different business sectors is another important factor to consider; for example, healthcare businesses take longer to establish and set up, while tech companies have a faster success rate due to their high customer engagement. E-commerce is another sector that can succeed quickly due to its higher social engagement ratio (Bento 2018).

Table 6 Common Business features and subfeatures with respect to level of difficulty (DL) to obtain, and its business relevance (BR) from low (+), medium (++), to high (+++)

Features	Sub features	Sub category of features	DL	BR
Administration	Organizational	Prior work experience	+	+
		Educational level	+	++
	Features	Use of technology	++	+++
		Salary of employee	+	+++
		managerial team size ratio	+	++
		No of managers	+	++
	Geographical	Location of the company	+	+++
	Location	Company size	+	+++
		Total population	++	+++
Demographics	Business Sector	Gender of employee per sector	++	+
		Age of the company	+	+++
		Number of employees per sector	++	+++
		Types of business sector	+	+++
	Balance sheet	Current Finances	+	+++
		Debt ratio	+	+++
		Number of equity	+	+++
	Cash Flow	Capital shares	++	+++
		Funds available	++	+++
		Total deposits	+++	++
		Total expenditure	++	+++
Financial	Income Statement	Total profit per sector	+	+++
Statements		Total earnings	+	+++
		Interest received	++	+++
		Total loss per year	++	+++
	Financial Ratio	Sum of expenses	++	++
		Profit to loss ratio	++	++
		Number of Investors	+	+++
		Revenue generated	+	+++
		Current assets/liability	+	+++
		Total equity, Total assets	+	+++
		Product quality	+++	+
Product	Product	Product quantity	++	++
	Information	Product distribution	+++	+
Features		Product price	+	++
		Market promotions	+++	+
		Product Sector, Product Category	+	+++
		Technology used in Product creation	+++	+++
	Social Media	Twitter/Facebook/Instagram Profile	+	+++
		Posts/tweets, followers	+	+++
Publicity	News	Articles of company	+	+++
		published reports about the products	+	++
	Web-based	Number of visits per page	+	++
		Customer reviews	+	++
		Questionnaire of the company	++	+++

3.2.3 Product features

Introducing a new product in the market requires careful planning and brainstorming new ideas and resources to ensure success. Products are essential for small and mid-size firms as they are the new source of revenue generation and developing opportunities for the employees to reach new customers in different market sectors (Florén et al. 2018). However, developing a new product requires careful consideration of the market capabilities and familiarity as it comes with a risk of failure if the market is still being determined. Therefore, it is important to know all the product and market distribution factors to achieve a successful business. Product quality is an important measure that ensures uniqueness and reduces competition in the market. To ensure business success, distribution and sales of the product require enough supply to fulfill the market needs. There should be the right amount of products with competitive prices to increase the profit margin of the sales in the company (Lian xxx). The success of the new product highly depends on the ideas and innovations used in creating a product, the technological capabilities in the business and the right type of market together contribute to the successful product and bring growth and resources into the business (Ernst 2002).

For companies providing services instead of products, product features may also refer features of the services, such as functionalities of software packages or review of restaurant in terms of locations, food quality, and customer/staff satisfaction (Li et al. 2023).

3.2.4 Financial statements

Financial statements are basic documents that reflect a company's financial status or performance. These statements are used by many financial institutions, government agencies and stockholders who can analyze and come up with a good idea about the future of the company's financial aspects. It gives information about the potential risks and challenges associated with investments, buying stocks from the market, granting bank loans for education, buying a property, or investing in a new business. Financial statements generally include a balance sheet, cash flow statements, income statements and business financial ratios. A detailed description of each financial statement is provided below.

- *Balance Sheet*—The balance sheet includes information about current value or finance of the business from a particular duration usually a quarter or a year. A balance sheet reveals whether the business has met its financial goal or how far they are from meeting its goal. A balance sheet usually contains three key components: financial assets, liability or equity (xxx).
- *Cash Flow* —Cash flow statements show the business's incoming and outgoing funds for a particular duration. It shows how the cash flow movement affects the balance sheet within the same duration. In a nutshell, cash flow statements reflect the business's financial health.
- *Income Statement*—The income statement is one of the most important parts of the financial statement, especially for investors looking deeply into the income to make necessary judgment about the business. An income statement is the profit and loss statement within a particular time duration. To evaluate whether the business has made profit, the income statement takes into consideration revenue, total expenses made and loss incurred over the duration of time.

- *Financial Ratios*—Financial ratios include current asset, liability, net income, total revenue and the cash flow to predict Profitability, Liquidity, Solvency and Efficiency of the business (Afolabi et al. 2019). They are financial indicators of the company evaluating the business performance. These indicators are derived from the financial statements generated from the business. Without financial indicators it would be impossible to predict the success or failure of the company (Xiang et al. 2012). This information is important for the decision makers to assess the company's profile and make decisions whether the company will be successful or fail. The below paragraph shows the formula to measure profitability, liquidity, efficiency and solvency of the business firm (Lukason and Käsper 2017).

$$\text{Profitability} = [\text{Net income/Revenue generated}] * 100\%$$

$$\text{Liquidity} = [\text{Current asset/ Current liability}] * 100\%$$

$$\text{Solvency} = [\text{Total equity / Total asset}] * 100\%$$

$$\text{Efficiency} = [\text{Revenue generated/ Total asset}] * 100\%$$

3.2.5 Publicity features

Business firms use social media platforms, like Twitter and Facebook, to promote their products or services to the public (Antretter et al. 2019). Social media tools are essential for start-up firms to market their products or spread awareness about their business to audience at relatively low cost (Saura et al. 2021). The information provided by firms and published on social media platforms attracts many investors who constantly seek new ventures with promising results. Social media features such as the number of followers, likes, comments and reviews provided by the users shield the light to estimate the popularity of the business (Jung and Jeong 2020).

Business firms not only look for social media platforms for the publicity of products or services in the market but also use news articles and websites to gain popularity from the public. Companies publish information about their products on websites or news articles, and the testimonials of satisfied customers are also disseminated through different channels (Xiang et al. 2012). It gives confidence to the customers to spend on the product and also provides reviews for potential buyers. Start-up firms and established businesses like Google or Facebook heavily rely on advertising to gain popularity (Sharchilev et al. 2018).

3.3 Market features

Market features are responsible for bringing success to the company by either investors or end users. With a good market, it is easier for a company to progress or grow in its field. An effective market can either make or break the company's existence (Kluwer 2020). Start-up businesses carefully evaluate all market features to anticipate the growth rate of the business.

Market growth and market sector are the two broad categories of market features. These two categories cover majority of the information about the company's marketing strategies and planning skills to gain profit in the business. A good marketing strategy is the utmost reason that brings revenue to the company. A company can strategize its marketing techniques to sell its products and gain profit in several ways. For example, identifying new customers who might be interested in the product as well as maintaining existing customers in the business by

Table 7 Market features and subfeatures with respect to level of difficulty (DL) to obtain, and its business relevance (BR) from low (+), medium (++), to high (+++)

Features	Sub Features	Sub Category of Features	DL	BR
Market Features	Market Sector	Creativity	+++	+
		Social media analytics	+++	+
		Investor sector	+	+++
		Sector wise growth	+++	++
		Uniqueness of the product per sector	+++	++
		Resource availability per sector	++	++
		Customer satisfaction	+	+++
	Service Market	Types of services offered	++	++
		Number of services sold	++	+++
		Review of services	+	+++
		Customer satisfaction based on services	++	+++
	Industrial Market	Market growth rate	+	+++
		Scope of market	++	+++
Market		Market competition	++	++
Growth		Marketing capabilities	+++	++
	Product Market	Product scope	++	++
		Product value	++	++
		Market share	+	++
		Market promotions	++	++
		Market target based on product	+	++

maintaining product quality (Thorleuchter et al. 2012), using social media platforms to market new products, analyzing the product with the correct market sector to gain utmost popularity and customer satisfaction (Azizi et al. 2009) are all important means of marketing strategy.

Table 7 summarizes market related features and its sub-features including market sectors and market growth *etc*.

All investment, business, and market features show some form of correlation if diving deep into the layers, as shown in Fig. 5. Technological features such as innovations or patents are highly correlated with product features in the Business section. Innovations bring new ideas and creations, which lead to developing new products in the market. Developing a product requires advertising and marketing strategies based on the product type and the market sector. Financial features in the investment section also strongly correlate to the business's financial features, which show the debts or cash flow ratio related to the investments or funding into the business.

Therefore, these features provide all aspects of analyzing business needs that are measurable and contributes to the growth of the business. With these features, we can support different types of companies and provide a well-fitted bias-free model to predict business success using machine learning algorithms.

4 Models for business success prediction

Business success prediction aims to analyze factors and features responsible for maintaining a profitable business and have a clear business target to produce measurable results. Analyzing business success is challenging due to unknown business data and fluctuations. Exploring business features and choosing a well-fitted model based on the company's different business angles and targets is important. In this section, we focus on describing several machine learning models categorized as Supervised or Unsupervised methods depending on the features' availability and the company's goal.

4.1 Supervised learning for business success prediction

In supervised learning method, the models are trained with a given label to predict the outcome of the data. Supervised machine learning has been used by many business companies to predict the outcome of the business. It first defines the target variable of the company with two possible outcomes success or failure. The data is split into subset of features represented by $\mathbf{x} \in \mathbb{R}^m$ (where m is the number of features, and $\mathbf{x}$ denotes a vector), the target variable is represented by $y \in \mathbb{R}$. The supervised learning model tries to predict the value of y for a given set of features $\mathbf{x}$.

Supervised learning models are divided into two subgroups: Regression and Classification, both have been used for predicting and forecasting business success. Based on the available label information models are divided further into three classes: Binary classification, Multi-class classification, and Continuous variable prediction, as shown in Table 8.

For example, Predicting business success can be a binary classification problem with label *successful* (1) or *unsuccessful* (0) or a multi-class classification problem with target variable containing more than two classes to predict stages of failure in private companies (Jones and Wang 2019). For a regression problem, the model uses a target with continuous variable (*e.g., Market trends*).

4.1.1 Binary classification

Binary classification is commonly used for predicting business success by many researchers using various machine learning algorithms described in the paragraph below.

- Support Vector Machine (SVM): An SVM algorithm is commonly used in binary classification tasks, where there are exactly two classes to predict. In a recent study (Li 2020), SVM is compared with Random Forest (RF) to classify business into two groups Fail ("closed") *vs.* Not-fail ("acquired" and "operating") using features from four major groups, including region, industry, funding, rounds, and name. Both SVM and RF show similar accuracy (around 88%), but their AUC values are very low with SVM being 0.51 and RF being 0.61. Because an AUC value with 0.5 implies a random classifier, this indicates that simple SVM is ineffective for business success prediction, possibly because of class imbalance, and features used in the study are less informative for classification.
- Logistic Regression (LR): Logistic regression is frequently used to predict "successful" or "unsuccessful", "failed" or "survived" for business prediction tasks. It is one of the mostly used algorithms due to its simplify and easy to interpret for business success

Table 8 Supervised learning models and their applications in business success prediction

Class Indicator	Business semantics	Business target	Features	Models
Binary	Success *vs*. Unsuccess	Acquired or Not Acquired Yeo and Grant (2018)	Business demographics and financial features	LR,SVM,gradient boosting
		IPO or Not Ünal (2019)	Business demographics and financial features	RF, extreme gradient boosting,LR
		Second round of funding or Not funded Li (2020)	Business financial features	SVM and RF
	Fail *vs*. Healthy	Survival or not survival Guerzoni et al. (2019)	Business demographics and investment technological features	SVM, LR, Naïve Bayes, ANN
		Successful or Fail Ravisankar et al. (2011) Ünal (2019) Krishna et al. (2016)	Business financial features	Probabilistic Neural Network, SVM, gradient boosting and Logistic Regression
Multi-class	Active, Failure, bankruptcy	Failure prediction Jones and Wang (2019) Sun et al. (2021)	Business financial features	Gradient boosting, Decision tree, Logistic Regression
	Risk,failure, bankrupt	Risk of bankruptcy Zhou et al. (2016)	Business financial and investment features	LR, NN, decision tree
	Acquired, funding,IPO	investment decision making in Acquired or ipo company Arroyo et al. (2019)	Business financial features	SVM,Decision Tree, GTB
	distress, Risk, failure	Financial risk in firms Cortes et al. (2007) Keasey and Watson (1991) Sun et al. (2014)	Business financial features	Adaboost, Decision tree,NN
	Risk,solvency,healthy	Financial stability Petropoulos et al. (2020)	Business financial features	Decision Tree,LR,SVM,RF
	Successful,Survival,Unsuccessful	Predicting survival of the company Guerzoni et al. (2021)	Business demographics and financial features	Naïve Bayes, Random Forest, Logistic regression
	Positive,negative,neutral tweets	Predict stock market trends Qasem et al. (2015)	Business publicity features	ANN,LR

Table 8 (continued)

Class Indicator	Business semantics	Business target	Features	Models
Continuous	Forecasting continuous growth	Predicting Profit Sharma et al. (2017)	Business financial features	Linear regression
	Customer churn Prediction	predicting customer behaviour churn Qureshi et al. (2013)	Market growth and business product features	Regression, decision tree and ANN
	Predicting growth of the company	sales prediction Cheriyan et al. (2018)	Business financial and product Features	Gradient Boost, Decision Tree
	High growth firms	Sales, profit and employment growth Zekić-Sušac et al. (2016) Delen et al. (2013)	Business financial features	Logistic regression ANN
	Growth range(positive, negative, zero)	sales growth Yeo and Grant (2018)	Business product features	Decision tree (CART)

prediction. A previous study (Żbikowski and Antosiuk 2021) has compared LR with SVM and XGBoost to predict successful and unsuccessful firms with target variable as 0 ('operating' or 'funding series b') and 1 ('acquired' and 'IPO'). By using an exhaustive grid search as hyper-parameter tuning, LR achieved an accuracy of 86%, however, the recall score of 0.21 was observed which improved to 0.34 when XGBoost classifier was used.

- Decision Tree (DT): Decision tree models are used for classification as well as regression tasks. Over the past few years, many financial institutions are constantly looking for simpler and effective models for prediction tasks related to financial sectors. A recent study demonstrated the use of decision tree algorithm for credit scoring application which is much simpler than previously used complex models that did not provide good results (Sohn and Kim 2012). The study distinguished start-up firms as either successful or default start-ups (*in terms of credit score*) by applying different input indicators (such as economic, financial, market indicators) to construct a decision tree model using a Classification and Regression Tree (CART) method. A classification accuracy of 74 % was achieved as compared to previously used LR model used for specific industry sectors, hence it would be difficult to generalize this model for other industry sectors.
- Naive Bayes: Naive Bayes is another classification problem which is based on the Bayes theorem which states that a posterior probability of an instance $\mathbf{x}$ belonging to class y, is defined by

$$P(y|\mathbf{x}) = \frac{P(\mathbf{x}|y)P(y)}{P(\mathbf{x})} = \frac{P(x_{i1}, \cdots, x_{im}|y)P(y)}{P(\mathbf{x})} = \frac{\prod_{j=1}^{m} P(x_{ij})P(y)}{P(\mathbf{x})} \tag{1}$$

  where $P(\mathbf{x}|y)$ is joint conditional probability of instance $\mathbf{x}$ with respect to the class y, and $P(y)$ is the prior probability of class y. According to the Naive Bayes assumption, all features are conditionally independent given the class label y, the joint conditional probability $P(\mathbf{x}|y)$ is simplified as the product of the conditional probability of all features $\prod_{j=1}^{m} P(x_{ij})$. Based on the study shown in Tomy and Pardede (2018) when predicting successful or failed firms from the list carefully selected features by extracting uncertainty factors from the original dataset, Naive Bayes algorithm have provided better accuracy of 77 % when compared to SVM and KNN.
- Artificial Neural Network (ANN): Artificial neural networks have been widely used for prediction of business success or failure in crowdfunding platforms. ANN uses a back propagation algorithm for the training process. The three layers in ANN: input layer, hidden layer(s), and output layer are responsible for carrying the information from one neuron to another and generate the desired output. In the recent study based on crowdfunding project (Alamsyah and Nugroho 2018), predictive modeling was performed to classify successful and unsuccessful projects using ANN. Different learning rates were applied out of which a learning rate of 0.2 with ANN gave an accuracy of 83 % which is beneficial for the investors to provide funding for the project.

Many researchers have used machine learning algorithms to predict business outcome of the startups or mid-size companies (Pasayat et al. 2020; Bento 2018; Dellermann et al. 2017; Ünal and Ceasu 2019). LR, SVM, and Gradient boosting have been commonly used (Żbikowski and Antosiuk 2021) to predict the success of the start-ups based on VC funding provided to the company. The main aim is to develop a bias-free prediction model so VCs and stakeholders can use it in real-world prediction scenarios without hesitation. A target variable is selected based on the completion of the second round of funding and labeled as "successful" because this marks the company's stability to generate enough profit in the future. In our previous study (Gangwani et al. 2023), we leverage triangle relationship between investors, business, and market to create new features predict business success (including acquisition and IPO) *vs.* failure. The results show that adding triangular relationship based features can indeed help improve the accuracy, compared to solutions using simple features alone.

Another definition of a successful start-up is based on the company's survival of the company. McKenzie et al. (2017) use three machine learning approaches, SVM, Least Absolute Shrinkage and Selection Operator (LASSO) and Boosted Regressor to predict the chances of survival based on sales and profit of the start-up companies. Böhm et al. (2017) suggests using techniques such as cluster analysis and SVM to predict the chances of survival.

Start-ups can also be measured based on innovations or the company's level of projects. Innovations can be a turning point as they can lead to reaching new heights and economic growth of the firms. Kinne and Lenz (2021) proposed a method to look at the innovation perspective of the business to predict the start-up success. A questionnaire-based survey was conducted on various firms to classify the data and label the firms' websites with new product innovations or non-innovations using a deep neural network architecture to capture innovations and predict the business outcome. Guerzoni et al. (2019) demonstrate that innovation increases the chances of survival in the company. Seven supervised learning approaches were used based on classification, Regression tree, Logistic Regression, Naive Bayes, and Artificial Neural Network (ANN) to predict the probability of survival in terms of innovations in the company.

4.1.2 Multi-class classification

Multi-class classification provides a different angle for entrepreneurs and investors to evaluate the business outcome. Many businesses do not gain profit over time, leading to the business's downfall. Various factors contribute to business failure, such as insufficient funds from the investors, poor marketing strategies, inability to compete with similar market etc. Over time, these factors become the sole reason for the company to fail in the market. Hence, this situation puts the company in danger of filing for bankruptcy. Multi-class algorithms help investors identify the risk of failure or default and give them an idea of where the company stands regarding profit.

The algorithm classifies business outcomes into multiple categories such as "risk", "failure", "survival" or "bankruptcy" rather than just classifying them as successful or not. This gives the investors a better chance to make wise decisions based on the companies' position and investment. Jones and Wang (2019) propose TreeNet method based on gradient boosting to perform classification tasks on private firms, which predicts the company's bankruptcy risk. As a result, their method can classify different measures of failure to analyze the company's risk before it files for bankruptcy. Arroyo et al. (2019) uses a time-aware analysis method to classify companies (as acquired, funding, or IPO) at their early stage to predict whether the company will be successful or not. The author uses a sliding window as a measure of time to give an estimate to the investors to decide on investment at the right stage of company's growth. The main aim of the investors is to invest when the company is at acquired or IPO stage. Machine learning algorithms have proven to be very useful and effective in predicting business outcomes. Due to the successful results delivered by machine learning models, many financial institutions such as corporate banks, credit scoring companies and insurance providers have moved from using statistical methods to machine learning models to predict financial crises or bankruptcy in their business.

4.1.3 Continuous variable prediction

In business prediction model, measuring the growth of the business is quantifiable. In order to effectively evaluate the business growth, a continuous evaluation process is needed to consider many factors and variables. The business's growth varies from company to company, and this is based on the defined target variable. For example, the growth can be measured in startups as sales profit or products sold. In mid-size companies, the growth can be measured based on the revenue generated, employment and customer service. Apart from these, various outside factors such as market, business environment, and product distribution should be considered when measuring the business's growth. Regression models in machine learning have proven to be very useful when predicting the growth of the business. Zekić-Šušac et al. (2016) use logistic regression and ANN to predict the company's growth using financial features from the agency. To predict a firm's growth, many interrelated features need to be considered, such as business financial features, geographical location and market portfolios *etc*. Despite identifying important features, researchers still struggle to quantify these features to predict the growth rate of the business accurately.

4.2 Unsupervised learning for business success prediction

Unsupervised machine learning model is also commonly used to analyze business related data, by finding hidden patterns or discover meaningful groups from a given dataset. One of most common advantage of unsupervised learning is that it doesn't rely on the labelled data to provide any information. There are four broad categories of unsupervised learning. Namely, clustering, association rule mining, outlier detection and dimensionality reduction. Table 9 summarizes different categories of unsupervised learning approaches and describes business implications and targets related to each category.

4.2.1 Clustering techniques

A clustering technique works by finding similar features which is useful for business prediction outcomes. For example, recommender system find customers sharing similar behaviors (e.g. similar purchases) and recommend items to them Li et al. (2015). Customer churn prediction Ullah et al. (2019) can be improved by clustering customer profiles in combination with the classification approach. By studying customers with similar behaviors, products with similar profiles, and companies with similar growth or failure history, clusters can help predict the chances of failure or survival of the business.

Many clustering methods exist for business data analysis, such as k-means clustering, partition based clustering, density based clustering, hierarchical clustering and model based clustering. Among them, due to its similarity and transparency, k-means clustering is most commonly used in business domains.

k-means clustering assigns n data points to k clusters, with the k value being specified beforehand. Each cluster is assigned a centroid during each iteration based on the distance of the data points to the centroid. Given a dataset with n observations $(\mathbf{x}_1, \cdots, \mathbf{x}_n)$ which

Table 9 Unsupervised learning models for business success prediction

Approaches	Business implication	Business target	Features	Models
Clustering	Group churn customers based on customer behaviour Ullah et al. (2019)	Churn prediction	Earning report	*k*-Means Clustering
	Clustering group of firms based on the time series analysis of financial loss Shah and Murtaza (2000)	Bankruptcy predicting	Financial ratios	Neural Network Clustering
	Clustering business models based on performance Böhm et al. (2017)	Success Prediction	Investors funding and revenue generated	*k*-means Clustering
	Clustering group of firms based on the financial situation such bankrupt or non bankrupt Lin et al. (2011)	Financial Crisis Prediction	Financial Statements	*k*-means and EM Clustering
	Group companies based on stock price for the investors to identify best companies Bini and Mathew (2016)	Stock Price increase for profit on returns	Sales and market features	*k*-Means, EM, Hierarchical and DBSCAN Clustering
	Companies were grouped based on growth and performance of the firms Zahay and Griffin (2010)	Firms sale and profit growth	Market sector and product and service price information	*k*-Means Clustering
	Group products and services based on consumers sentiments towards the product Halibas et al. (2018)	Business Growth	Social Media tweets and text	*k*-Means Clustering, sentiment analysis (NLP)

Table 9 (continued)

Approaches	Business implication	Business target	Features	Models
Association	Identify useful patterns in selection of customer needs and map them based on product information Jiao and Zhang (2005)	customer satisfaction	Product and sales information	Association rule mining (Apriori)
	Recommend products based on the market analysis and need of the customer Lin et al. (2002)	Profit sales	Market needs and product information	Association rule mining (Apriori)
	Discover relations between financial data and business operations Martin et al. (2011)	Predict bankruptcy	Financial Statements	Association Rule mining (Apriori, partition, FP-tree growth algorithms)
	Discover correlations between customers and market Aung et al. (2019)	Customer churn prediction	Sales and Revenue features	Aprori and FP growth
Outlier	Identify irregularities in business decision making Stamenković and Milanović (2014)	Business efficiency in competitive market	Product sales and market revenue	Fuzzy logic-based outlier detection
	Predict company's financial activity outliers Chen et al. (2007)	Predict financial crisis	Financial statements	Local outlier factor (LOF) outlier detection model
	Identify suspicious customers based on credit portfolio Figini et al. (2017)	Credit risk and financial solvency prediction	Investment reports and loan application statements	LOF outlier detection model

Table 9 (continued)

Approaches	Business implication	Business target	Features	Models
Dimen-sionality reduction	To identify similar patterns of financial features or variables in companies Wang and Wu (2017)	Business Failure Prediction	Financial Statements and cash flow	PCA, Kernal-based Self-organizing map (KFSOP)
	To analyze stock market movements patterns for future returns on invest-ment Zhong and Enke (2017)	Investment decision marking based on profit returns	Earning and Sales report	PCA and ANN
	To identify similar companies and map the companies together based on the financial aspects Lin et al. (2015)	Financial Distress	Product sales and financial statements	Isometric feature mapping (ISOMAP)
	To recognize risky end users and filtering bad credit user for the expansion of business Sivasankar et al. (2017)	Credit Risk Analysis	Business demographics and profit ratio report	PCA, LDA
	To recognize patterns determining financial aspect of healthy *vs.* under-perform firms Chakraborty and Sharma (2007)	Financial health prediction	Earning and sales report	PCA and SOFM

are assigned into k subsets $\mathcal{S} = \{S_1, \cdots, S_k\}$, the main objective of k-means clustering is to minimize the squared error function denoted by:

$$J(\mathcal{S}) = \arg\min_{\mathcal{S}} \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in S_i} ||\mathbf{x_j} - \boldsymbol{\mu}_i||^2 \tag{2}$$

where $\boldsymbol{\mu}_i$ denotes centroid of cluster i, which is calculated by using arithmetic mean of all data points in respective clustering.

Density-based clustering method is helpful in categorizing information together and filter outlier i.e. noise from the data. A study considered using clustering, diffusion theory, and density estimation for early prediction of sales data to analyze the success of a new product Garber et al. (2004). The key observation is that the clusters of data points, including the change of density, provide early warning signals to indicate the business's health.

Business Model DNA Böhm et al. (2017), similar to a human genome, describes the characteristics of the business process and uses a combination approach of SVM and k-means clustering to identify similar clusters based on different types of progress or growth in the business *(slow growth, fast growth etc.)* Using the combination of approaches, a better accuracy is achieved than previous studies based on the prediction of business success. Shah and Murtaza (2000) demonstrate using neural network with clustering techniques to predict bankruptcy in various firms. Three layers were used in the neural network architecture to predict bankruptcy. The first layer used financial ratio as an indicator to cluster the firms together, the second layer determined the learning process, which consisted time series data for predicting the trend of the financial status. The third layer consisted of two neurons, one classified bankrupt firms and other classified non-bankrupt firms.

4.2.2 Association rule mining

Association rule mining is used to analyze customer purchase behavior, which helps businesses make informed decisions about product placement, pricing, and promotional strategies. By analyzing patterns in customer purchase behavior, businesses can identify frequently purchased items together and use this information to design more effective product bundling and cross-selling strategies. For example in product portfolio identification (Jiao and Zhang 2005), recommendation (Lin et al. 2002), and market changing trend identification (Kaur and Kang 2016) etc.

Product portfolio identification (Jiao and Zhang 2005) is an important step that maps customer needs in the customer domain to functional requirements in the functional domain, as products are represented as a list of functional features and target values, which crucially decide the success and failure of the product.

The recommender system for e-commerce applications delivers personalized recommendations considering the similarities and dissimilarities of the customer's preferences. In the study (Lin et al. 2002), recommendation rules are mined for a specific customer to provide effective recommendations between customer and item rather than using traditional co-relation-based approach. An appropriate range is specified for calculating the $[minNumRule - maxNumRule]$ rules and a scoring threshold parameter for identifying ratings (*likes and dislikes*). Based on this rating which falls under the set of these rules, collaborative and target customers are identified to provide personalized recommendations to match customers' choices with items.

In financial sector of business domain, association rules help discover relations between business operations and financial health of the company. A set of rules are assigned to the financial data which discovers all combinations of business operations that put them in risk of being bankrupt (Martin et al. 2011). Apriori algorithm is used to find rules, in combination with financial domain ontology, to identify strong and weak points of company such as financial health or total debts to make decisions and helps them to strategize their plans and act accordingly.

Overall, association rule mining provides valuable insights into customer behavior, product portfolio, and financial analysis to help businesses make data-driven decisions about product placement, pricing, business operations, and promotional strategies, which can ultimately lead to increased sales and business success.

4.2.3 Outlier detection

In the business world, outliers often imply significant risks or values. Many financial sectors, such as banking industry, credit card sectors etc. have employed outlier detection models for identification and prediction of irregularities in the business domain. While outliers carry multiple forms, depending on the definition, local outliers are particularly useful because it helps identify samples not complying with others within a local neighborhood. Local Outlier Factor (LOF) compares the local density of the data point with the density of the neighboring data points. Previous study (Chen et al. 2007) demonstrated the use of LOF to detect inconsistencies or fraudulent activities in the banking industry to keep up with the reputation and provide customer satisfaction.

Given a data point p, its LOF score relies on several key concepts, including k-distance (p), k-Nearest Neighbor of (p), reachability distance (RD), and Local reachability distance (LRD).

For a dataset D and a positive integer k, the k-distance is the distance between the point $\mathbf{x}$ and its k^{th} nearest neighbor. Then $N_k(\mathbf{x})$ denotes the k nearest neighbors of $\mathbf{x}$, meaning that $N_k(x)$ includes all data points that lie in or on the circle of radius k-distance.

Reachability distance (RD) of a data point p to the data point $\mathbf{x}$, is defined as the maximum of k–distance of $\mathbf{x}$ and the distance between point p and $\mathbf{x}$, $d(p, \mathbf{x})$, which is calculated using Euclidean distance.

$$\mathrm{RD}(p, \mathbf{x}) = \max\{k\text{-distance}(\mathbf{x}), d(p, \mathbf{x})\} \tag{3}$$

For example, in Fig. 6, we take $k = 3$, the k nearest neighbors of instance $\mathbf{x}$ will be $N_k(\mathbf{x}) = \{p_1, p_2, p_3\}$. So the value of $||N_3(x)|| = 3$. If the distance between data point p_3 and

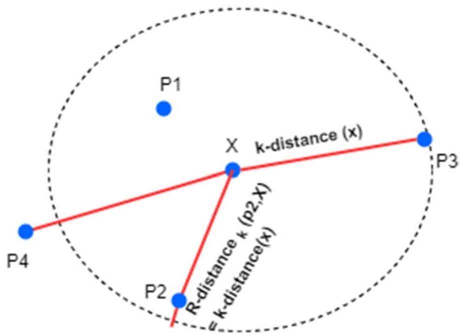

Fig. 6 An example demonstrating LOF and calculating reachability distance of data points

data point x is smaller than the k-distance ($\mathbf{x}$), the reachability distance of data point p_3 is replaced by k-distance($\mathbf{x}$). For data point p_4, its distance between p_4 and $\mathbf{x}$ is greater than the k-distance ($\mathbf{x}$), the reachability distance of p_4 is the original distance of p_4 and $\mathbf{x}$.

Combining k-distance($\mathbf{x}$) and reachability distance RD(p, $\mathbf{x}$), density-based method uses two parameters for calculating density: 1) MinPts, which specifies the minimum number of points; 2) volume. These two parameters determine the density threshold θ to detect the outliers. For keeping our calculation of LOF simple, we remove the MinPts parameter as the measure of volume and with this knowledge, we calculate the local reachability density of data point (p).

$$LRD_k(p) = \frac{1}{\sum_{\mathbf{x}\in N_k(p)} \frac{RD(p,\mathbf{x})}{||N_k(p)||}} \tag{4}$$

The LRD of each point is compared with the average LRD of k neighbors. LOF is the ratio of average LRD of k neighbors of (p) to the LRD of (p), denoted by

$$LOF_k(p) = \frac{\sum x \in N_k(P)^{Lrd_k(p)}}{||N_k(p)||} \times \frac{1}{LRD_k(p)} \tag{5}$$

Overall, outlier detection can provide businesses with valuable insights into their data, helping them to improve their operations and manage risks by helping them detect the risks such that businesses can take necessary actions to mitigate the impact of these risks and make better decisions, and ultimately, improve their business strategies.

4.2.4 Dimensionality reduction

Dimensionality reduction is a popular technique used in various areas of data analysis, including business prediction. It is used to simplify complex datasets by reducing the number of features (i.e., dimensions) while retaining the most important information.

Wang and Wu (2017) proposed a two-stage ensemble approach to improve the performance of the business failure prediction using feature selection to eliminate redundant data having little or no information about the financial features. The final subset includes carefully selected financial indicators that cover information about healthy and failed firms. Three manifold learning algorithms (ISOMAP, Liner Embedding (LE) and Local Linear Embedding (LLE)) were applied to select different subset of features and compare their performance with PCA which enhanced the model's performance.

Another study aimed to predict business bankruptcy using financial ratios (Tsai 2009). PCA was used to reduce the dimensionality of the data and identify the most important financial ratios for predicting bankruptcy. The results showed that using dimensionality reduction techniques improved the accuracy of the bankruptcy prediction model. Specifically, PCA reduced the dimensionality of the data from 14 financial ratios to 5 principal components, accounting for 91% of the total variance.

Recent studies (Sivasankar et al. 2017; Rtayli and Enneya 2019) have demonstrated feature selection and feature extraction (PCA and LDA) techniques for credit risk or fraud detection. By combining Random Forest and feature-filtering, the research (Rtayli and Enneya 2019) proposes to detect credit card fraud, with relevant features selected from the financial dataset, using a feature importance scoring calculated using RF classifier with *gini index* to construct a decision tree that determines the final class in each tree. The representation of *gini index* at node n, $g(n)$ which measure the impurity, is given as:

$$g(n) = 1 - \sum_{i=1}^{c} p_i^2 \tag{6}$$

where, p_i calculates the probability of class in the given branch and a 0 *gini index* score means that all elments belong to one particular class.

The studies highlight the importance of dimensionality reduction in business prediction, as it can help identify the most important variables for predicting a target variable and improve the accuracy of prediction models (Table 10).

4.3 Deep learning for business success prediction

Recently, many studies have applied deep learning methods such as CNN (convolutional neural network), LSTM (long short-term memory), and DNN (deep neural networks) for business prediction. The main advantage of using these methods is that they learn news features independently without strong domain expertise or requiring hand-crafted features. Since business success prediction includes textual data collected from social media websites, news headlines, finance and banking industry, it is necessary to represent them into vectors and feed them into machine learning models for better prediction. This section examines different approaches to selecting textual features and covert them into vectors using deep learning methods.

In order to convert a sentence or textual data, such as news articles and social media websites, for sentiment analysis, NLP packages like Word2vec and Doc2vec are used. In a recent deep learning-based business failure prediction (BFP) method (Borchert et al. 2022), word embedding is used to convert textual data to numeric form which is then used as an input to the convolutional layer. For example, a sentence in the form of $w = [w_1, w_2, ..., w_m]$ with length m, using the word embedding to stack them in the form of a matrix represented as $N = (n_1^T, n_2^T, ..., n_m^T)^T \in \mathbb{R}^{m*l}$ where n_i is the embedded word representation of a sentence w_i. Hence, each word in a sentence can be represented as a vector $n \in \mathbb{R}^l$ where l is the number of dimensions.

Using deep learning methods for handling textual features have shown tremendous improvements when compared to traditional machine learning methods specifically in topics related to bankruptcy and stock price prediction (Oncharoen and Vateekul 2018; Qu et al. 2019). Classical deep learning models, such as LSTM and BERT (Li et al. 2023) have proven to be effective in working with financial and customer data. A study about restaurant survival prediction (Li et al. 2023) categories customers' online review into five categories: location, tastiness, price, service, and atmosphere, by using pretrained language model (BERT) and aspect-based sentiment analysis. Their results show that aspect-based sentiment, compared to overall review sentiment, can improve the prediction performance of restaurant survival. In Ding et al. (2014), a series of events representing news articles and headlines are extracted as a tuple denoting an event which consists of three values (Action, Actor and Object). Based on the articles, three values are extracted which together become an event. For example in the headline "Apple launches a new iPhone" where Actor = "Apple", Action = "Launches" and Object = "iPhone". These extracted tuples are transformed into word vectors using pre-trained GloVe method (Pennington et al. 2014) that extracts only meaningful words. The vector is fed into the LSTM network as an input. The output from the LSTM network is fed into hidden layers connected with two perceptrons, which uses softmax activation function that serves as an output to the prediction model.

Table 10 A summary of data and resources for business success prediction

Dataset	Description	Records/Data information
Crunchbase Żbikowski and Antosiuk (2021); Ross et al. (2021); Dellermann et al. (2017); Pan et al. (2018); Liang and Yuan (2016) https://data.crunchbase.com/docs/daily-csv-export	It is a platform that provides business information such as investors, founders, funding rounds, organization information, employee details etc. of public and private firms	Structured data containing daily CSV snapshot of 50 records for categories such as (company, people,organization,funding round,acquisition)
TechCrunch Krishna et al. (2016) Xiang et al. (2012) Sharchilev et al. (2018) https://data.world/aurielle/techcrunch-startup	An online newspaper platform containing information about startups and cutting-edge technological firms available in the market. It is a platform for latest market trends and online advertisement about the companies and investors	Data contains 212 records of news articles from(company, people and products) in a csv format
Kickstarter Greenberg et al. (2013) Jhaveri et al. (2019) Li et al. (2016) https://github.com/sdevalapurkar/kickstarter-prediction	One of the leading crowdfunding platforms which provides information about funding, investments and creativity	Contains structured data about funded projects with over 187399 records generated in csv file
USPTO Wei et al. (2008) Leydesdorff et al. (2014) Kim and Lee (2015) Lee et al. (2009) https://www.uspto.gov/ip-policy/economic-research/research-datasets/patent-assignment	An agency that issues patents to investors for their businesses and new inventions. It also issues trademark registration for the Intellectual Property(IP)	Contains structured data of publicly available patent of two types(Patent assignment data(with 8.97 million patents and Patent examination research data(PATEx) with 16.5 million patents)
AngelList Beckwith (2016) Zhang et al. (2017) https://kyang01.github.io/startup-analysis/	It is one of the famous crowdfunding platforms for startups and potential investors for funding	Dataset can be scraped either by using default API or github to gather information about tweets and start-up companies https://github.com/kyang01/startup-analysis/tree/master/data

The main motivation behind using deep learning methods is that it carries out feature learning on the original data, such as texts or images, to find important factors for business success prediction. Deep learning provides a new perspective to researchers in business prediction as it can help diversify data sources without involving strong domain knowledge and heavy data prepossessing.

4.4 Survival analysis for business success prediction

For most business success prediction, the models are based on one-off success/failure prediction using a snapshot of feature input. Survival analysis, which concerns the time between existing and subsequent events, provides an alternative to assess the risk of business success/failure considering the business course as a time series events (Wang et al. 2022; Zhou et al. 2022). The rational is that business entering bankruptcy often behave differently during a series of events happened before. In this context, survival analysis employs a parametric model assuming some relationships, such as Kaplan-Meier or Cox Regression, between the hazard event (or survival) and the set (vector) of explanatory variables (X). For example, by modeling business behaviors and recurrence of financial distress as a survival analysis model, a stratified conditional intensity model (Zhou et al. 2022), defined in Eq. (7) is proposed to model instantaneous risk for a firm i to have the k^{th} distress at time t, given that it has survived till time t.

$$h_{ik}(t|X) = \delta_{ik}(t)h_{0k}(t)\exp(\beta X + \alpha X^{Int} + \gamma Z) \tag{7}$$

In Eq. (7), $\delta_{ik}(t) \in \{0, 1\}$ is the modified risk-set indicator denoting the happening of the k^{th} event at time t. $h_{0k}(t)$ denotes baseline hazard function for the k^{th} recurrence of financial distress. α, β, γ are regression parameters, and X denotes features representing company (such as features like institutional variables, financial ratios, market-based variables and macro economic conditions. X^{Int} denotes interaction between features X, and Z denotes correlation between different distress episodes.

In addition to modeling business success, survival analysis is also frequently used in business intelligence to predict a customer's churn time (Ahn et al. 2020).

4.5 AI and generative models in business success

Recent advancement in AI, especially large language models like ChatGPT, has set off tsunami of discussions about embracing AI innovations to the industry 4.0, business operation and success (Javaid et al. 2023; Brem et al. 2023; Jorzik et al. 2023; Raj et al. 2023). Traditionally, AI is considered enabling technologies to help identify correlations, features, and similarities in large amounts of data for many business operation. Nevertheless, as AI is evolving to be much more powerful and intelligent, it is now reshaping customer interactions and shaking services and processes (Brem et al. 2023), because AI-enable tools provide quick, informative, and more natural responses. Another key advantage of AI tools, like ChatGPT, lies in its fine-tuning ability which is able to tailor models responses to specific industry domains or niches. This helps customize products/services for enhanced customer engagement, and more positive experience for the customers (Raj et al. 2023).

In addition to leverage AI technologies at product/service levels, a recent study (Javaid et al. 2023), conducted through the interview of 47 top managements and AI specialists, proposes a top management (TM) competencies model for business executives to enforce AI thinking in mindset, knowledge, navigability, leadership, decision-making ability.

4.6 Summary of business success prediction

Business prediction methods summarize the importance of selecting the best method based on the business target. For example, unsupervised learning methods are most commonly used for analyzing financial health and marketing products. Supervised learning methods utilize different business features to predict business success or startup survival. Recently, deep learning methods have proven to be more efficient for business failure prediction using financial and historical data. Our study outlines typical methods for business prediction using important and measurable criteria. Based on the data availability and business target, researchers and entrepreneurs can effectively utilize different methods provided to create a successful business model.

To further compare different methods for business success prediction, in terms of different business targets and their interrelated features, Fig. 7 highlights the IBM features and the models used for business prediction. The methods are aligned and positioned based on their focus and targets. For example, the methods that use market and business features would be placed along the edges connecting Market and Business.

To compare the niche of three methods used for modeling and predicting business success, we summarize them in Table 11 concerning available features, advantages, and disadvantages.

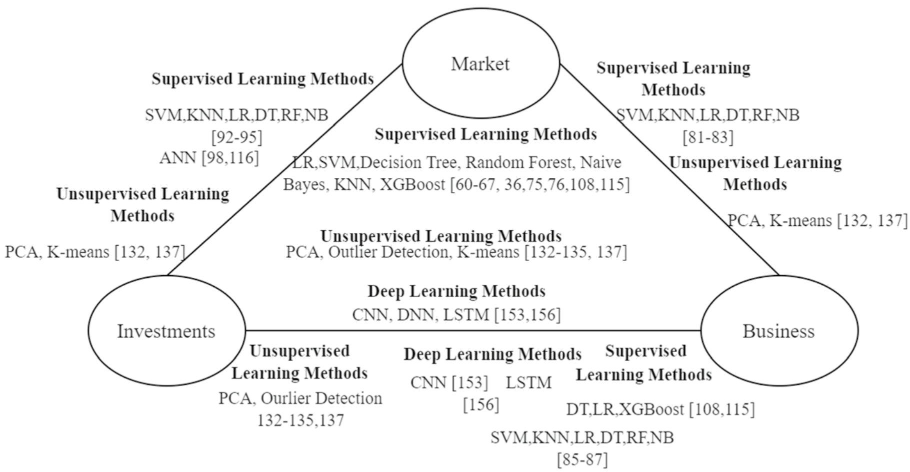

Fig. 7 A summary of business success prediction methods and their focus with respect to the IBM triangle for business modeling

Table 11 Summary of business prediction methods with IBM features

Methods	Advantages	Disadvantage	Commonly available features	May not be available
Supervised ML Models Ravisankar et al. (2011); Wei et al. (2008); Saura et al. (2021); Bento (2018); Shah and Murtaza (2000); Sun et al. (2021); Chandler and Hanks (1994)	Good control over the training data	Tends to overfit high dimensional data	Investment—Financial features,founder info, technology Business-Demographics, Financial statements,Product info Market-Market sector/type	Investment—HR,Innovations Business-Administrative, Publicity, Web-based reviews Market - Market growth
Unsupervised ML Models Edgett and Parkinson (1994); Lussier and Halabi (2010); Heidari et al. (2021); Shah and Murtaza (2000)	Can detect hidden patterns which may not be visible to humans	Does not guarantee usefulness of results	Investment—Financial features Business-financial statements,product info Market -Market sector/type	Business—Administrative, Publicity, Web-based reviews Market - Market growth
Deep Learning Models Vochozka et al. (2020); Allu and Padmanabhuni (2022); Rule and Ambady (2008)	Does not require fine-tuning, capability to use textual features	Requires more memory to train the model	Investment—Financial features Business-financial statements,publicity, customer review Market -Market sector	Business—Administrative features

5 Data and performance metrics

We describe and list various publicly available datasets collected from different platforms for modeling and predicting business outcomes. Meanwhile, we also review several performance metrics commonly used to validate business success.

5.1 Data and resources

Table 10 summarizes datasets and provides a detailed description. Many researchers use Crunchbase dataset (Żbikowski and Antosiuk 2021; Ross et al. 2021; Xiang et al. 2012; Arroyo et al. 2019; Yuxian and Yuan 2013) which is a publicly available dataset that provides detailed information about public and private companies. It contains important information about companies, such as the investors, founders, acquisition details, funding rounds and employee details, that are needed to predict business success depends on the company's target or goal of the prediction.

Various startup companies use TechCrunch as a popular newspaper in the U.S. platform dedicated to profiling start-ups that provide up-to-date information about the latest products and trends in the market (Sharchilev et al. 2018; Xiang et al. 2012; Bargagli Stoffi et al. 2020). Start-ups rely on such information to achieve a specific milestone or growth in the business. The data collected from these platforms provide the basis for the prediction of success in terms of the company's M &A, IPO or financial survival.

United States Patent and Trademark Office (USPTO) is another source that provides information about patents and trademarks. Companies holding Intellectual Property (IP) rights and patents mark the technological innovations and creativity which distinguish them from other companies (Lee et al. 2009; Kim and Lee 2015). The data collected from such platforms highlights the capacity of the company to grow and become an IPO which is an important indicator to predict the business success for the investors and the stakeholders.

There are also various other resources such as stock market dataset and resources gathered from financial banks which provides information about the financial aspects of the companies, the profit and loss statements, sales and cash flow information *etc*. Ravisankar et al. (2011). This type of information is very useful for small and mid-size companies to predict success or failure of the firms.

Social media platforms such as Twitter or Facebook have also been used to collect important information about companies, such as their followers, products sold, number of tweets, reviews about the products etc. Antretter et al. (2019); Saura et al. (2021). These publicity features highlight some important questions such as "How famous is the company?" and "How many customers buy the products sold by these firms?" (Yuxian and Yuan 2013; Liang and Yuan 2016). Other crowdfunding platforms such as AngelList, a U.S.-based website contemplated connecting stakeholders with investors for the purpose of funding (Beckwith 2016). All these resources provide a better outlook for investors to decide whether to invest in a company. It is also useful for the stakeholders to know the position of the company and anticipated growth and progress of the company.

5.2 Performance metrics

Evaluating the performance of the business is done by measuring factors contributing to the business's success or failure. We evaluate business metrics in terms of loss and performance in two ways: the performance in terms of machine learning models and the performance in terms of business interests. A confusion matrix-based performance metrics also highlights the actual positives and actual negatives from the predicted values.

5.2.1 Confusion matrix based performance metrics

Confusion matrix is the most common metric used to evaluate the model's performance for binary as well as multi-class classification task. For business success prediction, *TruePositive*, *TtrueNegative* and *FalsePositive*, *FalseNegative* are used to evaluate the actual and the predicted values (Turkmen 2021) as shown in Eqs. (9) and (10). The confusion matrix also helps to denote the Type1 and Type 2 errors statistics which is useful to evaluate business prediction models as same metrics can be used in different models for comparison between two approaches for the same problem.

5.2.2 Performance metrics in terms of learning models

During the learning, the evaluation of the predictive modeling is done in two steps: 1) Loss function, and 2) Performance metrics. In loss function the model is evaluated by using the in-sample-loss-minimization function.

$$\operatorname{argmin} \sum_{i=1}^{N} \ell\left(f\left(\mathbf{x}_i\right), y_i\right) \text{over} f(\cdot) \in F \text{ s.t. } R(f(\cdot)) \leq c \tag{8}$$

where $\sum_{i=1}^{N} \ell\left(f\left(\mathbf{x}_i\right), y_i\right)$ calculates the mean squared error of prediction, known as the loss function, which is to be minimized, $f\left(\mathbf{x}_i\right)$ denotes predicted values and y_i are the actual values, $f(\cdot) \in F$ is denoted as the function class of the algorithm, and $R(f(\cdot))$ is known as the complexity function which is expected to be less than a constant value $c \in \mathbb{R}$.

From a performance metrics point of view, the model's performance is evaluated based on the algorithm chosen and the type of metrics it supports.

1. *F-score (or F1-score)*: *F*-score metric is specifically used for imbalanced datasets where one class is more dominant than others. It is calculated using both precision and recall score as shown in Eqs. (11) and (12). For binary classification task, *F*-score is useful to validate the model performance in terms of both classes.
2. *AUC*: Another important performance metric used for both binary and multi-class classification is Area Under the receiver operating characteristic Curve (AUC). It is one of the most frequently used metric for classification problem since it is independent from the use of false positive/negative cost.
3. *Kolmogorov-Smirnov (KS)*: KS chart is a measure of degree of separation between two classes (positive and negative). Based on the business target, for example customer churn prediction or market segment analysis, KS chart is a very useful metric to predict the probability of two classes so that the business can target specific set of customers.
4. *Kappa score*: Cohen's Kappa score is used to measure the probability of agreement (p_r) between two classes on the scale of 0–1 as shown in Eqs. (13). It can be used in binary as well as multi-class classification problem. For example in project evaluation where the outcome of success is divided into number of success indicators (Wohlin et al. 2000). Kappa score is more useful than accuracy when dealing with imbalanced data.

$$\text{Sensitivity/Recall} = \frac{TP}{TP + FN} \tag{9}$$

$$\text{Specificity (True Negative)} = \frac{TN}{TN + FP} \tag{10}$$

$$\text{Precision (True Positive)} = \frac{TP}{TP + FN} \tag{11}$$

$$\text{F1-Score} = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{12}$$

$$\text{Kappa score} = \frac{Accuracy - p_r}{1 - p_r} \tag{13}$$

5.2.3 Performance metrics in terms of business interests

Some common measures to evaluate the business performance and interests are summarized as follows:

1. *Return on Equity (ROE)*: ROE is a financial measure calculated by dividing a company's net income by its total equity. Investors and stakeholders are more interested in knowing the returns on their investments so that they can evaluate the performance of the company and decide their next course of action. ROE provides an easily understandable metrics without digging much further into the finances and investments.
2. *Debt Ratio*: Debt ratio is the percentage of total debt to total asset ratio. In an uncertain market, the risk of investments increases the company's exposure to unexpected downturns. Hence, the Debt ratio is a good evaluation metric for the stakeholders to anticipate the shift in the market and measure where the company stands in terms of profit or loss.
3. *Stocks Buyout*: Buyout of stocks is a scenario where investors acquire the original or failed company agreeing on a lower percentage of stocks buyout such that the failed company can exit without any debt and investors can take advantage of rectifying the market with ease. In order to measure the business loss, the buyout stock price should be closer to the market value.
4. *Liquidity Measure*: As the name suggests, liquidity measure is nothing but the amount of available cash which can be used for business purposes within a short duration of time. Measuring liquid cash flow is an important criterion to know whether the company can withstand ups and downs in the market within a short span of time. Having low cash flow is one of the indicators to measure business loss.
5. *Total Revenue*: Total revenue is the earnings incurred by the company after selling products/goods and services. These metrics determine whether the company reached its goal either annually or semi-annually determining profit or loss in the market.
6. *Profit*: Total profit is calculated by dividing the income by total expense. A quarterly evaluation of these metrics helps the company to stay on top and change its goals or marketing to attain a higher number.

6 Conclusion and future works

This study fills the research gap concerning business cycle fluctuations and the key factors contributing to business success, as well as explores computational methods for predicting business success. So far, very few studies have explored the factors affecting business growth and links between business-related theories and important features for predicting success. Without proper knowledge and facts, predicting business success is limited to specific use cases or business sectors. In this paper, we first address theories related to business fluctuations and leverage the theories to identify features and factors relevant to the business success. An IBM triangle framework is proposed to highlight three different angles of business features and demonstrate their relation in business operation. The framework provides flexibility for researchers to extend the architecture with additional features and factors for different types of businesses. Based on the gathered knowledge, we review different business prediction methods using machine learning and deep learning models and compare popular algorithms regarding their business semantics, target, features, and models. Our review surveys different machine learning and deep learning methods based on their applicability and business targets. The study also provides in-depth knowledge to help understand key components of machine learning approaches for business modeling and prediction. The categorization and summary of critical factors and features responsible for business modeling help researchers comprehensively understand the existing methods and provide substantial resources for entrepreneurs and investors to broaden machine learning and deep learning models to different domains.

Our study is limited to measurable features for the prediction of business success as we focus on data-driven strategies such that the companies can set measurable goals based on current and historical data. Future study can focus on several key aspects of using such features for business prediction. First, business often requires transparent and interpretable features/models, a study of actionable factors for business prediction is important for stakeholders to put such computational models into real-world usages. Second, the reliability of each algorithm or model depends on several factors, such as business size, location, target, and data availability, and business requires quantifiable reliability, in addition to accuracy/error, for effective decision support. For example, existing study (Ross et al. 2021) uses financial features such as debt-to-equity ratio, returns on investment, and profit margin to determine financial health of venture capitalist firm and its growth potential for future success using a combination of deep learning and machine learning boosting algorithms. Similarly, another study (Krishna et al. 2016) focuses on analyzing business outcomes using financial features and market-based analysis such as competitive product pricing, and product distribution using supervised machine learning algorithms for predicting private firms' success. A study of reliable and safe learning models for business success is also an important topic for future study.

Acknowledgements This research is sponsored in part by the U.S. National Science Foundation through Grant Nos. IIS-2236579, IIS-2302786 and by NSF Industry-University Cooperative Research Center for Advanced Knowledge Enablement (CAKE), FAU site.

References

Abdullah M (2021) The implication of machine learning for financial solvency prediction: an empirical analysis on public listed companies of bangladesh. J Asian Bus Econ Stud 28(4):303–320

Acciarini C, Cappa F, Boccardelli P, Oriani R (2023) How can organizations leverage big data to innovate their business models? a systematic literature review. Technovation 123:102713

Afolabi I, Ifunaya TC, Ojo FG, Moses C (2019) A model for business success prediction using machine learning algorithms. J Phys: Conf Ser 1299:012050

Ahn J, Hwang J, Kim D, Choi H, Kang S (2020) A survey on churn analysis in various business domains. IEEE Access 8:220816–220839

Aksoy H et al (2017) How do innovation culture, marketing innovation and product innovation affect the market performance of small and medium-sized enterprises (smes). Technol Soc 51(4):133–141

Akyildirim E, Goncu A, Sensoy A (2021) Prediction of cryptocurrency returns using machine learning. Ann Oper Res 297(1):3–36

Alamsyah A, Nugroho TBA (2018) Predictive modelling for startup and investor relationship based on crowdfunding platform data. In: Journal of Physics: Conference Series, vol. 971, pp 012002. IOP Publishing

Al-Henzab J, Tarhini A, Obeidat BY et al (2018) The associations among market orientation, technology orientation, entrepreneurial orientation and organizational performance. An Intl. J, Benchmarking

Ali-Yrkkö J, Hyytinen A, Pajarinen M (2005) Does patenting increase the probability of being acquired? evidence from cross-border and domestic acquisitions. Appl Financ Econ 15(14):1007–1017

Allu R, Padmanabhuni VNR (2022) Predicting the success rate of a start-up using lstm with a swish activation function. J Control Decis 9(3):355–363

Altman EI (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. J Finan 23(4):589–609

Antretter T, Blohm I, Grichnik D, Wincent J (2019) Predicting new venture survival: a twitter-based machine learning approach to measuring online legitimacy. J Bus Venturing Insights 11:00109

Arroyo J, Corea F, Jimenez-Diaz G, Recio-Garcia JA (2019) Assessment of machine learning performance for decision support in venture capital investments. IEEE Access 7:124233–124243

Ashton WB, Sen RK (1988) Using patent information in technology business planning-i. Research-Technology Mgmt 31(6):42–46

Aung MM, Han TT, Ko SM (2019) Customer churn prediction using association rule mining. J Trend Sci R &D 3(5):1886–1890

Azizi S, Movahed SA, Khah MH (2009) The effect of marketing strategy and marketing capability on business performance case study: Iran's medical equipment sector. J Med Market 9(4):309–317

Bargagli Stoffi F, Riccaboni M, Rungi A (2020) Machine learning for zombie hunting. firms' failures & financial constraints. ML for Zombie Hunting. Firms' Failures & Financial Constraints (April 28, 2020)

Barro RJ (1990) The stock market and investment. Rev Financ Stud 3(1):115–131

Beckwith JJ (2016) Predicting success in equity crowdfunding

Bento FRdSR (2018) Predicting start-up success with machine learning. PhD thesis

Bini B, Mathew T (2016) Clustering & regression techniques for stock prediction. Procedia Technol 24:1248–1255

Boasson V, MacPherson A (2001) The role of geographic location in the financial and innovation performance of publicly traded pharmaceutical companies: empirical evidence from the untied states. Environ Plan A 33(8):1431–1444

Böhm M, Weking J, Fortunat F, Müller S, Welpe I, Krcmar H (2017) The business model dna: Towards an approach for predicting business model success

Borchert P, Coussement K, De Caigny A, De Weerdt J (2022) Extending business failure prediction models with textual website content using deep learning. Eur J Operat Res 306(1):348–357

Bouri E, Lau CKM, Lucey B, Roubaud D (2019) Trading volume and the predictability of return and volatility in the cryptocurrency market. Financ Res Lett 29:340–346

Breitzman A, Thomas P (2002) Using patent citation analysis to target/value m &a candidates. Research-Tech Mgmt 45(5):28–36

Breitzman A, Thomas P, Cheney M (2002) Technological powerhouse or diluted competence: techniques for assessing mergers via patent analysis. R &D Manag 32(1):1–10

Brem A, Giones F, Werle M (2023) The ai digital revolution in innovation: a conceptual framework of artificial intelligence technologies for the management of innovation. IEEE Trans Eng Mgmt 70(2):770–776

Brik AB, Rettab B, Mellahi K (2011) Market orientation, corporate social responsibility & business performance. J Business Ethics 99:307–324

Brüderl J, Preisendörfer P, Ziegler R (1992) Survival chances of newly founded business organizations. American Soc Rev 57:227–242

Cardless G Five types of financial statements (June2021). https://gocardless.com/en-us/guides/posts/types-of-financial-statements/

Cerra V, Fatas A, Saxena SC (2020) Hysteresis and business cycles. IMF Working Papers **2020**(073)

Chakraborty S, Sharma SK (2007) Prediction of corporate financial health by artificial neural network. J Elec Financ 1(4):442–459

Chandler GN, Hanks SH (1994) Market attractiveness, resource-based capabilities, venture strategies, and venture performance. J Bus Ventur 9(4):331–349

Chang S-B, Lai K-K, Chang S-M (2009) Exploring technology diffusion and classification of business methods: Using the patent citation network. Technol Forecast Soc Change 76(1):107–117

Chawla SK, Pullig C, Alexander FD (1997) Critical success factors from an organizational life cycle perspective: perceptions of small business owners from different business environments. J Bus Entrep 9(1):47

Chen M-C, Wang R-J, Chen A-P (2007) An empirical study for the detection of corporate financial anomaly using outlier mining techniques. In: Intl. Conf. on Convergence Information Technology, pp. 612–617

Cheriyan S, Ibrahim S, Mohanan S, Treesa S (2018) Intelligent sales prediction using machine learning techniques. In: Intl Conf on Computing, Electronics & Communications Engineering (iCCECE). IEEE

Chin JT (2020) Location choice of new business establishments: Understanding the local context and neighborhood conditions in the united states. Sustainability 12(2):501

Choi J, Jeong B, Yoon J, Coh B-Y, Lee J-M (2020) A novel approach to evaluating the business potential of intellectual properties: a machine learning-based predictive analysis of patent lifetime. Comput Ind Eng 145:106544

Cortes EA, Martinez MG, Rubio NG (2007) Multiclass corporate failure prediction by adaboost m1. Int Adv Econ Res 13(3):301–312

Crain M (2014) Financial markets & online advertising: reevaluating the dotcom investment bubble. Info Comm Soc 17(3):371–384

Delen D, Kuzey C, Uyar A (2013) Measuring firm performance using financial ratios: a decision tree approach. Expert Syst Appl 40(10):3970–3983

Dellermann D, Lipusch N, Ebel P, Popp KM, Leimeister JM (2017) Finding the unicorn: Predicting early stage startup success through a hybrid intelligence method

Deng Z, Lev B, Narin F (1999) Science and technology as predictors of stock performance. Financ Anal J 55(3):20–32

Devi SS, Radhika Y (2018) A survey on machine learning and statistical techniques in bankruptcy prediction. J ML Comp 8(2):133–139

Dibb S (1998) Market segmentation: strategies for success. Marketing Intelligence & Planning

Ding X, Zhang Y, Liu T, Duan J (2014) Using structured events to predict stock price movement: An empirical investigation. In: intl conf on empirical methods in natural language proc., pp. 1415–1425

Drábek J, Lorincová S, Javorčíková J (2017) Investing in human capital as a key factor for the development of enterprises. Issues Human Resour Manag 1(1):113–136

Edgett S, Parkinson S (1994) The development of new financial services: identifying determinants of success and failure. Int J Serv Indus Manag 5(4):24–38

Editors H (2017) The Best-Performing CEOs in the World 2017. https://hbr.org/2017/11/the-best-performing-ceos-in-the-world-2017?autocomplete=true

Ernst H (2001) Patent applications and subsequent changes of performance: evidence from time-series cross-section analyses on the firm level. Res Policy 30(1):143–157

Ernst H (2002) Success factors of new product development: a review of the empirical literature. Intl J Manag Rev 4(1):1–40

Feldman SP (1988) How organizational culture can affect innovation. Organ Dyn 17(1):57–68

Figini S, Bonelli F, Giovannini E (2017) Solvency prediction for small and medium enterprises in banking. Decision Support Sys 102:91–97

Florén H, Frishammar J, Parida V, Wincent J (2018) Critical success factors in early new product development: a review and a conceptual model. Intl Entrepr Manag J 14(2):411–427

Four ways HR can contribute to maximum ROI. https://www.pockethrms.com/blog/4-ways-hr-can-contribute-to-maximize-roi/

Fuertes-Callén Y, Cuellar-Fernández B, Serrano-Cinca C (2020) Predicting startup survival using first years financial statements. J Small Bus Manag 60(6):1314–1350

Gangwani P, Perez-Pons A, Bhardwaj T, Upadhyay H, Joshi S, Lagos L (2021) Securing environmental iot data using masked authentication messaging protocol in a dag-based blockchain: Iota tangle. Future Internet 13(12):312

Gangwani D, Zhu X, Furht B (2023) Exploring investor-business-market interplay for business success prediction. J Big Data 10(1):1–28

Gangwani P, Soni J, Upadhyay H, Joshi S (2020) A deep learning approach for modeling of geothermal energy prediction. International J Comput Sci Inf Secur (IJCSIS) **18**(1)

Garber T, Goldenberg J, Libai B, Muller E (2004) From density to destiny: using spatial dimension of sales data for early prediction of new product success. Mark Sci 23(3):419–428

Gorgievski MJ, Ascalon ME, Stephan U (2011) Small business owners' success criteria, a values approach to personal differences. J Small Bus Manage 49(2):207–232

Greenberg MD, Pardo B, Hariharan K, Gerber E (2013) Crowdfunding support tools: predicting success & failure. In: CHI'13 Ext. Abs. on Human Factors in Computing Systems, pp. 1815–1820

Greenley GE (1995) Market orientation & company performance: empirical evidence from UK companies. British J mgmt 6(1):1–13

Greenwood J, Jovanovic B (1990) Financial development, growth, and the distribution of income. J Politi Econ 98(5):1076–1107

Guerzoni M, Nava CR, Nuccio M (2021) Start-ups survival through a crisis combining machine learning with econometrics to measure innovation. Econo Innov New Technol 30(5):468–493

Guerzoni M, Nava CR, Nuccio M (2019) The survival of start-ups in time of crisis. a machine learning approach to measure innovation. arXiv preprint arXiv:1911.01073

Halibas AS, Shaffi AS, Mohamed MAKV (2018) Application of text classification & clustering of twitter data for business analytics. In: 2018 Majan International Conference (MIC), pp. 1–7. IEEE

Hawkins R (2004) Looking beyond the dot com bubble: exploring the form and function of business models in the electronic marketplace, 65–81

Heidari M, Zad S, Rafatirad S (2021) Ensemble of supervised and unsupervised learning models to predict a profitable business decision. In: IEEE Intl. IOT, Electronics & Mechatronics Conference, pp. 1–6

Hou J-J, Chien Y-T (2010) The effect of market knowledge management competence on business performance: a dynamic capabilities perspective. Int J Electr Bus Manag 8(2):96

Hunt SD (2013) A general theory of business marketing: R-a theory, alderson, the isbm framework, and the imp theoretical structure. Ind Mark Manag 42(3):283–293. https://doi.org/10.1016/j.indmarman.2013.02.002

Innovation and Intellectual Property. https://www.wipo.int/ip-outreach/en/ipday/2017/innovation_and_intellectual_property.html

Iyer P, Davari A, Zolfagharian M, Paswan A (2019) Market orientation, positioning strategy and brand performance. Ind Mark Manage 81:16–29

Jaafari A (2001) Management of risks, uncertainties and opportunities on projects: time for a fundamental shift. Int J Proj Manag 19(2):89–101

Javaid M, Haleem A, Singh RP (2023) A study on chatgpt for industry 4.0: background, potentials, challenges, and eventualities. J Econ Technol 1:127–143

Jhaveri S, Khedkar I, Kantharia Y, Jaswal S (2019) Success prediction using random forest, catboost, xgboost and adaboost for kickstarter campaigns. In: 3rd Intl Conf on Computing Methodologies & Communication, pp. 1170–1173. IEEE

Jiao J, Zhang Y (2005) Product portfolio identification based on association rule mining. Comput Aided Des 37(2):149–172

Jones S, Wang T (2019) Predicting private company failure: a multi-class analysis. J Int Financ Markets Inst Money 61:161–188

Jones S, Wang T (2019) Predicting private company failure: A multi-class analysis. J Int Finan Markets Inst Money 61:161–188. https://doi.org/10.1016/j.intfin.2019.03.004

Jorzik P, Yigit A, Kanbach DK, Kraus S, Dabić M (2023) Artificial intelligence-enabled business model innovation: Competencies and roles of top management. IEEE Trans Eng Mgmt, pp 1–13

Jung SH, Jeong YJ (2020) Twitter data analytical methodology development for prediction of start-up firms' social media marketing level. Technol Soc 63:101409

Justiniano A, Primiceri GE, Tambalotti A (2010) Investment shocks and business cycles. J Monet Econ 57(2):132–145. https://doi.org/10.1016/j.jmoneco.2009.12.008

Kakati M (2003) Success criteria in high-tech new ventures. Technovation 23(5):447–457

Kaur M, Kang S (2016) Market basket analysis: Identify the changing trends of market data using association rule mining. Procedia Comput Sci 85:78–85

Keasey K, Watson R (1991) Financial distress prediction models: a review of their usefulness 1. British J Manag 2(2):89–102

KENTON W (2022) Fortune 100 Definition, Requirements, and Top Companies

Kerr WR, Lerner J, Schoar A (2010) The consequences of entrepreneurial finance: a regression discontinuity analysis. Technical report, National Bureau of Economic Research

Kim J, Lee S (2015) Patent databases for innovation studies: a comparative analysis of uspto, epo, jpo and kipo. Technol Forecasting Soc Change 92:332–345

Kim B, Kim H, Jeon Y (2018) Critical success factors of a design startup business. Sustainability 10(9):2981

Kinne J, Lenz D (2021) Predicting innovative firms using web mining and deep learning. PLoS ONE 16(4):0249071

Kleinaltenkamp M, Jacob F (2002) German approaches to business-to-business marketing theory: origins and structure. J Bus Res 55(2):149–155

Kluwer W (2020) Business success depends upon successful marketing. https://www.wolterskluwer.com/en/expert-insights/business-success-depends-upon-successful-marketing

Kozielski R (2019) Determinants of smes business success-emerging market perspective. Intl J Organiz Anal 27(2):322–336

Krishna A, Agrawal A, Choudhary A (2016) Predicting the outcome of startups: less failure, more success. In: 2016 IEEE 16th Intl. Conf. on Data Mining Workshops (ICDMW), pp. 798–805. IEEE

Lee A (2013) Welcome to the unicorn club: Learning from billion-dollar startups

Lee S, Yoon B, Lee C, Park J (2009) Business planning based on technological capabilities: patent analysis for technology-driven roadmapping. Technol Forecast Soc Chang 76(6):769–786

Lee K, Park I, Yoon B (2016) An approach for r &d partner selection in alliances between large companies, and small and medium enterprises (smes): application of bayesian network and patent analysis. Sustainability 8(2):117

Lee R, Lee J-H, Garrett TC (2019) Synergy effects of innovation on firm performance. J Bus Res 99:507–515

Leydesdorff L, Kushnir D, Rafols I (2014) Interactive overlay maps for us patent (uspto) data based on international patent classification (ipc). Scientometrics 98(3):1583–1599

Li J (2020) Prediction of the success of startup companies based on support vector machine and random forset. In: 2020 2nd Intl. Workshop on AI and Education, pp. 5–11

Li B, Zhu X, Li R, Zhang C (2015) Rating knowledge sharing in cross-domain collaborative filtering. IEEE Trans Cybern 45(5):1068–1082

Li H, Yu BXB, Li G, Gao H (2023) Restaurant survival prediction using customer-generated content: an aspect-based sentiment analysis of online reviews. Tour Manage 96:104707

Lian I Eight stages of new product. https://smallbusiness.chron.com/eight-stages-new-product-55433.html#content

Liang YE, Yuan S-TD (2016) Predicting investor funding behavior using crunchbase social network features. Internet Res 26(1):74–100

Lin W, Alvarez SA, Ruiz C (2002) Efficient adaptive-support association rule mining for recommender systems. Data Mining Knowl Discov 6(1):83–105

Lin W-Y, Hu Y-H, Tsai C-F (2011) Machine learning in financial crisis prediction: a survey. IEEE Trans Syst Cybern, Part C Appl Rev 42(4):421–436

Lin F, Yeh C-C, Lee M-Y (2015) Integrating nonlinear dimensionality reduction with random forests for financial distress prediction. J Testing Eval 43(3):645–653

Li Y, Rakesh V, Reddy CK (2016) Project success prediction in crowdfunding environments. In: Proceedings of 9th ACM intl conf on web search and data mining, pp. 247–256

Lukason O, Käsper K (2017) Failure prediction of government funded start-up firms. Invest Mgmt Financ Innov 14(2):296–306

Lussier RN, Halabi CE (2010) A three-country comparison of the business success versus failure prediction model. J Small Bus Manage 48(3):360–377

MacMillan IC, Zemann L, Subbanarasimha P (1987) Criteria distinguishing successful from unsuccessful ventures in the venture screening process. J Bus Venturing 2(2):123–137

Mankiw NG (1995) Real business cycles: a new Keynesian perspective. J Econ Perspect 3(3):79–90

Marr B (2019) The 10 Best Examples Of How Companies Use Artificial Intelligence In Practice. https://www.forbes.com/sites/bernardmarr/2019/12/09/the-10-best-examples-of-how-companies-use-artificial-intelligence-in-practice/?sh=1aece2a67978
Martin A, Manjula M, Venkatesan DVP (2011) A business intelligence model to predict bankruptcy using financial domain ontology with association rule mining algorithm. arXiv preprint arXiv:1109.1087
Mason CM, Harrison RT (2002) Barriers to investment in the informal venture capital sector. Entrep Reg Dev 14(3):271–287
McKenzie, D. David J. & Sansone: Man vs. machine in predicting successful entrepreneurs: evidence from a business plan competition in nigeria (2017)
Mokhber M, Khairuzzaman W, Vakilbashi A (2018) Leadership and innovation: the moderator role of organization support for innovative behaviors. J Mgmt Org 24(1):108–128
Morgan NA (2012) Marketing and business performance. J Acad Mark Sci 40:102–119
Oncharoen P, Vateekul P (2018) Deep learning for stock market prediction using event embedding and technical indicators. In: Intl. Conf. on advanced informatics: concept theory and app., pp. 19–24
Pan C, Gao Y, Luo Y (2018) Machine learning prediction of companies 'business success. CS229: Machine Learning, Stanford University
Pasayat AK, Bhowmick B, Roy R (2020) Factors responsible for the success of a start-up: a meta-analytic approach. IEEE Trans Eng Manag 70(1):342–352
Pennington J, Socher R, Manning CD (2014) Glove: Global vectors for word representation. In: Proc. of the 2014 Conf. on empirical methods in natural language processing (EMNLP), pp. 1532–1543
Perboli G, Arabnezhad E (2021) A machine learning-based dss for mid and long-term company crisis prediction. Expert Sys Appl 174:114758. https://doi.org/10.1016/j.eswa.2021.114758
Petropoulos A, Siakoulis V, Stavroulakis E, Vlachogiannakis NE (2020) Predicting bank insolvencies using machine learning techniques. Intl J Forecasting 36(3):1092–1113
Qasem M, Thulasiram R, Thulasiram P (2015) Twitter sentiment classification using machine learning techniques for stock markets. In: Intl. Conf. on Advances in Comput., Communications & Info., pp. 834–840
Qu Y, Quan P, Lei M, Shi Y (2019) Review of bankruptcy prediction using machine learning and deep learning techniques. Procedia Comput Sci 162:895–899
Qureshi SA, Rehman AS, Qamar AM, Kamal A, Rehman A (2013) Telecommunication subscribers' churn prediction model using machine learning. In: Eighth international conference on digital information management (ICDIM 2013), pp. 131–136. IEEE
Rai A, Patnayakuni R, Patnayakuni N (1997) Technology investment and business performance. Commun ACM 40(7):89–97
Raj R, Singh A, Kumar V, Verma P (2023) Analyzing the potential benefits and use cases of chatgpt as a tool for improving the efficiency and effectiveness of business operations. BenchCouncil Trans Benchmarks Standards Eval 3(3):100140
Ravisankar P, Ravi V, Rao GR, Bose I (2011) Detection of financial statement fraud and feature selection using data mining techniques. Decis Support Syst 50(2):491–500
Ross G, Das S, Sciro D, Raza H (2021) Capitalvx: a machine learning model for startup selection & exit prediction. J Financ Data Sci 7:94–114. https://doi.org/10.1016/j.jfds.2021.04.001
Rtayli N, Enneya N (2019) Credit card risk detection based on feature-filter and fraud identification. In: 2019 third international conference on intelligent computing in data sciences (ICDS), pp. 1–8. IEEE
Rule NO, Ambady N (2008) The face of success: inferences from chief executive officers' appearance predict company profits. Psychol Sci 19(2):109–111
Samuelson PA (1939) Interactions between the multiplier analysis and the principle of acceleration. Rev Econ Stat 21(2):75–78
Sandvik IL, Sandvik K (2003) The impact of market orientation on product innovativeness and business performance. Int J Res Market 20(4):355–376
Santisteban J, Mauricio D (2017) Systematic literature review of critical success factors of information technology startups. Acad Entrep J 23(2):1–23
Santisteban J, Mauricio D, Cachay O (2021) Critical success factors for technology-based startups. Int J Entrep Small Bus 42(4):397–421
Saura JR, Reyes-Menéndez A, deMatos N, Correia MB (2021) Identifying startups business opportunities from ugc on twitter chatting: an exploratory analysis. J Theor Electr Comm 16(6):1929–1944
Schumpeter JA (2017) The theory of economic development: An inquiry into profits, capita i, credit, interest, and the business cycle
Shah JR, Murtaza MB (2000) A neural network based clustering procedure for bankruptcy prediction. Am Bus Rev 18(2):80

Sharchilev B, Roizner M, Rumyantsev A, Ozornin D, Serdyukov P, de Rijke M (2018) Web-based startup success prediction. In: Proceedings 27th ACM Intl. Conf. on Info. & Knowledge Mgmt, pp. 2283–2291
Sharma A, Bhuriya D, Singh U (2017) Survey of stock market prediction using machine learning approach. Intl Confl Electr Commun Aerosp Technol 2:506–509
Sivasankar E, Selvi C, Mala C (2017) A study of dimensionality reduction techniques with machine learning methods for credit risk prediction. In: Behera HS, Mohapatra DP (eds) Comput Intell Data Mining. Springer, Singapore, pp 65–76
Sohn SY, Kim JW (2012) Decision tree-based technology credit scoring for start-up firms: Korean case. Expert Syst Appl 39(4):4007–4012
Solesvik M, Gulbrandsen M (2013) Partner selection for open innovation. Technol Innov Manag Rev 3(4):6–11
Song Y-G, Cao Q-L, Zhang C (2018) Towards a new approach to predict business performance using machine learning. Cogn Syst Res 52:1004–1012
Stamenković M, Milanović MM (2014) Outlier detection in function of quality improvement of business decisions. In: Proceedings of the international scientific conference–enterprises in hardship: economics, managerial & juridical perspectives, pp. 173–184
Stuart R, Abetti PA (1987) Start-up ventures: towards the prediction of initial success. J Bus Ventur 2(3):215–230
Sun J, Li H, Huang Q-H, He K-Y (2014) Predicting financial distress & corporate failure: a review from state-of-the-art definition modeling sampling & featuring approaches. Knwl-Based Sys 57:41–56
Sun J, Fujita H, Zheng Y, Ai W (2021) Multi-class financial distress prediction based on support vector machines integrated with the decomposition and fusion methods. Inf Sci 559:153–170
Teichert T, Ernst H (1999) Assessment of r &d collaboration by patent data. In: PICMET'99: Portland International Conference on Management of Engineering and Technology. Proceedings Vol-1: Book of Summaries (IEEE Cat. No. 99CH36310), pp. 78–86. IEEE
Thorleuchter D, Van den Poel D, Prinzie A (2012) Analyzing existing customers' websites to improve the customer acquisition process as well as the profitability prediction in b-to-b marketing. Expert Syst Appl 39(3):2597–2605
Tomy S, Pardede E (2018) From uncertainties to successful start ups: a data analytic approach to predict success in technological entrepreneurship. Sustainability 10(3):602
Tsai C-F (2009) Feature selection in bankruptcy prediction. Knowl-Based Syst 22(2):120–127
Tucker R (2021) Innovation Can Absolutely. Positively Be Measured, Four Tips From The Pros For Doing It Right
Turkmen E (2021) Deep learning based methods for processing data in telemarketing-success prediction. In: Intl Conf on Intelligent Communication Tech. & Virtual Mobile Networks, pp. 1161–1166
Ullah I, Raza B, Malik AK, Imran M, Islam SU, Kim SW (2019) A churn prediction model using random forest: analysis of machine learning techniques for churn prediction and factor identification in telecom sector. IEEE Access 7:60134–60149
Ünal C (2019) Searching for a unicorn: A machine learning approach towards startup success prediction. Master's thesis, Humboldt-Universität zu Berlin
Ünal C, Ceasu I (2019) machine learning approach towards startup success prediction. Technical report, IRTG 1792 Discussion Paper
Vochozka M, Vrbka J, Suler P (2020) Bankruptcy or success? the effective prediction of a company's financial development using lstm. Sustainability 12(18):7529
Vui CS, Soon GK, On CK, Alfred R, Anthony P (2013) A review of stock market prediction with artificial neural network(ann). In: IEEE Intl. Conf. on Control Sys., Comp. & Eng., pp. 477–482. IEEE
Wan X (2010) A literature review on the relationship between foreign direct investment & economic growth. Int Bus Res 3(1):52
Wang L, Wu C (2017) Business failure prediction based on two-stage selective ensemble with manifold learning algorithm & kernel-based fuzzy self-organizing map. Knowl-Based Syst 121:99–110
Wang Z, Jiang C, Zhao H (2022) Know where to invest: platform risk evaluation in online lending. Info Sys Res 33(3):765–783
Wei C-P, Jiang Y-S, Yang C-S (2008) Patent analysis for supporting merger and acquisition (m &a) prediction: A data mining approach. In: Workshop on E-Business, pp. 187–200. Springer
Weill P (1992) The relationship between investment in information technology and firm performance: a study of the valve manufacturing sector. Inf Syst Res 3(4):307–333
Why now is the perfect time to invest in HR (2020). https://www.forbes.com/sites/sap/2020/09/01/why-now-is-the-perfect-time-to-invest-in-human-resources/?sh=37bfed2d56b5

Wohlin C, Von Mayrhauser A, Höst M, Regnell B (2000) Subjective evaluation as a tool for learning from software project success. Inf Softw Technol 42(14):983–992

Xiang G, Zheng Z, Wen M, Hong J, Rose C, Liu C (2012) A supervised approach to predict company acquisition with factual and topic features using profiles and news articles on techcrunch. In: Sixth Intl AAAI Conference on Weblogs and Social Media

Yazdipour R, Constand R (2010) Predicting firm failure: a behavioral finance perspective. J Entrep Financ 14(3):90–104

Yeo B, Grant D (2018) Predicting service industry performance using decision tree analysis. Intl J Inf Manag 38(1):288–300

Yu P-F, Huang F-M, Yang C, Liu Y-H, Li Z-Y, Tsai C-H (2018) Prediction of crowdfunding project success with deep learning. In: 15th Intl. Conf. on E-business Engineering, pp. 1–8

Yuan X, Hou F, Cai X (2020) How do patent assets affect firm performance? from the perspective of industrial difference. Technol Anal Strateg Manag 33(8):943–956

Yuxian EL, Yuan S-TD (2013) Investors are social animals: Predicting investor behavior using social network features via supervised learning approach. Workshop on Mining and Learning with Graphs

Zahay D, Griffin A (2010) Marketing strategy selection, marketing metrics, & firm performance. J Bus Indus Market 25(2):84–93

Zane M (2023) How Many New Businesses Started In 2022. https://www.zippia.com/advice/how-many-new-businesses-started/

Żbikowski K, Antosiuk P (2021) A machine learning, bias-free approach for predicting business success using Crunchbase data. Info Process Mgmt 58(4):102555

Zekić-Sušac M, Šarlija N, Has A, Bilandžić A (2016) Predicting company growth using logistic regression and neural networks. Croatian Oper Res Rev 7(2):229–248

Zhang Q, Ye T, Essaidi M, Agarwal S, Liu V, Loo BT (2017) Predicting startup crowdfunding success through longitudinal social engagement analysis. In: Proceedings of 2017 ACM on Conf on information & knowledge management, pp. 1937–1946

Zhong X, Enke D (2017) Forecasting daily stock market return using dimensionality reduction. Expert Sys App 67:126–139

Zhou L, Tam KP, Fujita H (2016) Predicting the listing status of Chinese listed companies with multi-class classification models. Inf Sci 328:222–236

Zhou F, Fu L, Li Z, Xu J (2022) The recurrence of financial distress: a survival analysis. Intl J Forecast 38(3):1100–1115