Financial Analytics

Mahmut Sami Sivri, Abdullah Emin Kazdaloglu, Emre Ari, Hidayet Beyhan, and Alp Ustundag

1 Introduction

Financial analytics is the application of data science methods and techniques in finance domain. Data analytics has a significant and rising role in helping companies reduce risk and make more efficient financial decisions. AI-based real-time risk detection systems are used to assess risks across the company. Financial statements are automatically examined, including balance sheets, income statements and cash flows, and early warning systems are established. Credit allocation decisions are made using sophisticated machine learning models that estimate how likely a consumer is to repay a loan. Customers receive smart recommendations from financial organizations based on their banking or investment preferences. Financial firms use algorithmic trading and sophisticated mathematical models to develop new trading techniques. Fraudulent transactions and money laundering activities can be detected using anomaly detection algorithms. Robo-advisors provide automated and algorithm-driven financial planning services to their customers without human intervention. In this chapter, four cases are presented regarding financial statement analysis, credit risk management and investment analytics.

M. S. Sivri · A. E. Kazdaloglu · E. Ari · A. Ustundag (✉)
Department of Industrial Engineering, Faculty of Management, Istanbul Technical University, 34367 Macka, Istanbul, Turkey
e-mail: ustundaga@itu.edu.tr

H. Beyhan
Department of Management Engineering, Faculty of Management, Istanbul Technical University, 34367 Macka, Istanbul, Turkey
e-mail: beyhan17@itu.edu.tr

A. Ustundag et al. (eds.), *Business Analytics for Professionals*, Springer Series in Advanced Manufacturing, https://doi.org/10.1007/978-3-030-93823-9_13

2 Financial Statements, Ratio Analytics and Bankruptcy Prediction

In its simplest terms, financial analysis is a process that is generally carried out to generate long-term plans and to evaluate the financial performance, convenience and sustainability of the company in the context of the company's sector, economic conjuncture and financial policies. This process could be adapted not only to the company context but also to projects, investments, budgets and other finance-based transactions. If the financial analysis is done externally, the answer can be sought whether any business is profitable and stable enough to guarantee the return of the monetary investment made in it. If done internally, it enables managers and financial analysts within the company to draw inferences from the company's past financial situation. It helps them make their decisions and recommendations about the future of the company. In this way, financial analysis reveals the strengths and weaknesses of the company that affects its competitive position among other companies, and it is in a principal position for the development of the company [1]. When both ways are evaluated in detail, financial analysis, in short, expresses the company's capacity to carry out its operational processes profitably, to fulfil its obligations and to generate sufficient cash for the opportunities that may arise instantly in the market. At the same time, it should not be forgotten that the company is financially responsible for issues such as paying interest, meeting the principal debt and paying dividends to its investors, who try to maximize their earnings by minimizing their risks [2].

In other words, financial analysis is a general term that expresses the ability to make business and investment-based decisions using financial data. This analysis is performed in two different ways under the headings of technical and fundamental analysis. Technical analysis is statistical trend analysis regarding the company's past data such as share price changes and sales volume. Technical analysis is based on understanding the behaviour pattern of investors in the market rather than looking for the main reasons underlying the share price. In this way, it is tried to make it possible to predict the fluctuations in the shares more easily according to the actions of the market participants. On the contrary, fundamental analysis provides a perspective on the fundamental value of the company by using the financial statements of the company and the ratios derived from these statements. Typically, analysts make this analysis by considering macroeconomic factors such as government policies, financial regulations, microeconomics such as internal policies and management of the company, and environmental factors, including the internal balances of the company's industry [3]. In this way, it creates a foresight to value security. Generally, fundamental analysis is carried out in two ways, and these ways are qualitative and quantitative methods. Qualitative methods include issues such as brand awareness, patents and technology, while quantitative methods are carried out using the three primary expressions used in financial analysis (balance sheet, profit & loss statement and cash flow statement). Thus, companies use these tables both for their internal operations and for the convenience and transparency of reporting to their investors. All three tables are interrelated and shed light on the company's activities, status and financial performance [4]. In the light of all this information, it is obvious that financial statements are of great importance to determine the status of companies.

2.1 Problem Definition

Fundamental financial analysis is one of the most important financial building blocks for all companies and is carried out by examining the financial reports of companies. Fundamental analysis includes many distinct steps in order to present an entire picture of the company's performance. The starting point for this application is the company's financial statements, specifically the income statement (profit & loss statement), balance sheet and cash flow statement. The income statement is used to analyse the company's performance in terms of sales, revenue, business profitability and future cash flow projections over a certain period of time and is usually the starting point of financial statement analysis. The balance sheet shows the financial position of the company at a specific point in time. The balance sheet includes the company's resource (assets) and capital (liability and equity) information. The cash flow statement, on the other hand, shows the operational, financial and investment cash flow, source and use of the company in a certain time period by highlighting the liquidity, solvency and financial flexibility of the company [5]. In the light of all these financial statements, it is possible to calculate the performance or value of any company and to make trend and ratio analysis for each statement with various performance metrics.

However, despite all the benefits that financial statements provide for the company's situation, there are certain limitations as well. First of all, companies' financial statements are mostly unique and the wording in the statements varies from company to company. For instance, a company in the music industry expresses its sales under the name 'Total Revenue', while another company in the retail industry shows this expression as 'Total Net Sales'. Another problem is that different companies collect various financial information differently to elicit cumulative sums. Because of such differences, making comparisons between companies stands out as one of the constraints of financial statements. Another problem is that financial information can be misleading due to the fact that financial statements provide data for a specific time period and effects such as seasonality. The third problem is that there are many financial ratios. Although each of these ratios has a healthy proportional range, it is not possible to evaluate several of them together manually. The fourth problem can be expressed as the fact that the financial statements do not provide enough information about the future legitimacy of the companies, since they contain historical data. The firm's viability and legitimacy are of great importance to the firm's creditors and investors, and therefore, it is critical to assess the firm's probable bankruptcy. In order to eliminate all these problems, it is clear that companies should seek robust and dynamic solutions when evaluating their financial statements.

With this point of view, it is critical to render financial statements with dense information more understandable, to visualize important parameters in all three financial statements and to calculate important ratios. Moreover, it is necessary to develop a system supported by machine learning methods that will evaluate whether the company will go bankrupt in the light of the financial information and ratios in all three financial statements. In this context, answers will be sought to the following main issues:

- What factors and financial information should be considered when analysing the company's historical data?
- What are the changes in the profitability, efficiency, liquidity and solvency of the company over a certain period of time?
- How to visualize important financial information?
- What kind of model should be developed in order to make the comparison of financial ratios of different companies meaningful and to evaluate the health of companies for the future? Which ratios and/or metrics should be included in this model?

2.2 Case Study: A Technology Company

Founded in the autumn months of 2000, X company is a multinational technology company with its headquarters in Y country and more than 1 million employees worldwide, serving both online and offline platforms. In 2020, more than 1 billion unique users visited the website of company X, whose focus is electronic commerce. In 2020, more than 1 billion products were sold through the websites of company X, and the total annual revenue from these sales (online + offline) reached the $400 billion-mark. On a monthly basis, the website of company X receives approximately 15 million clicks and over 50 thousand transactions are made. At the same time, company X has 150 offline stores worldwide and more than 100 million products are sold in these stores on an annual basis.

In this case study, the financial statements of Company X from the first quarter of 2001 to the last quarter of 2020 will be our primary focus. Although there are fluctuations from time to time in the 20-year period, it can be indicated that Company X is in a serious upward trend in terms of financial information, such as sales, net income, assets, capital and net cash flow. Company X's income statement includes financial items for which cumulative sums can be calculated, such as net sales, gross profit, total operating expenses, operating income, profit before tax and net profit. Assets in the balance sheet consist of current and non-current assets. Its liabilities, on the other hand, include current and non-current liabilities and include long-term liabilities among them. In addition, the total equity of the stakeholders is also shown. The cash flow statement of company X is basically divided into three as operating, investment and financial cash flow. As a whole, company X's income statement, balance sheet and cash flow statement are described item by item and all financial statements of Company X have the following assumptions and operational and financial environment:

- The data in the financial statements of Company *X* was not defected intentionally or unintentionally, and these data were audited by an independent auditor and an audit opinion was presented for the accuracy of the financial statements.
- Financial statements of *X* company online and offline stores are evaluated together.
- It is assumed that government policies, macroeconomic changes and potential competitors do not have an impact on Company *X*'s financial statements and will not have any future effect.
- The companies in the data set to be used to predict bankruptcy are not affiliated with Company *X* in any way, and there is no correlation between their financial information.
- All aspects of technical analysis, such as company X's share price, are not covered by this case study.

The IT department of *X* company wants to develop a system by directly assisting its creditors and investors externally, and its financial analysts and senior management internally. In this context, the balance sheet, income statement and cash flow statement of *X* company will be analysed in terms of 20-financial year period, and this analysis will be supported by dynamic graphics and financial ratio calculations will be made for the financial health of the company. Thus, financial statements will be saved from their complex structure and will be made easier to comprehend. In addition, a bankruptcy prediction model will be developed using machine learning methods and data from more than 6800 companies in country *Z*. In this way, the data obtained by *X* company using the current financial statement analysis methods will be tested with a more realistic model, and thus, it will be possible to have clearer information about the future and legitimacy of the company. Hence, in the context of this case study, first of all, comprehensive financial statement and ratio analysis of *X* company will be made and then the bankruptcy prediction of *X* company will be made with the help of the data containing financial information of more than 6800 companies used in machine learning methods. A foresight will be created regarding the future financial situation of the company. The data sets containing the financial statements of *X* company and the financial data of 6800 companies are given in the data set folder of the chapter.

2.3 Model

In this case study, the model is examined under two main headings. First of all, a structure will be established in which the general trend of *X* company is analysed, financial ratio analysis is made, and all findings are supported by graphics. Then, a comprehensive bankruptcy prediction model will be developed with machine learning algorithms and the bankruptcy prediction of company *X* will be made thanks to the findings obtained from the first subsection.

2.4 *Fundamental and Ratio Analysis*

In this part of the model, the three financial statements will be carefully examined, and the quarterly data will be expressed as fiscal years, the cumulative totals of the financial statements will be calculated, and thus, the completed financial statements will be visualized. Then, the important financial items in all three financial statements will be graphed. Finally, financial ratios will be calculated by making the necessary calculations and the ratios will be visualized on the basis of categories.

2.4.1 Profit and Loss Statement Analysis

Profit & loss statement is a financial report used by companies for internal control and planning, and externally for stakeholders to evaluate the company's financial performance [6]. This report summarizes the financial information of companies such as revenue, loss, expense, profit/loss for a certain period of time (i.e. quarterly, fiscal year). Profit & loss statement, which is also referred to as a profit statement or a statement of activities in some sources, shows the ability of companies to maximize their profits by increasing their revenue and reducing their expenses. Essentially, it most quickly reflects the financial return of how much profit or loss companies generate with the work they do.

2.4.2 Balance Sheet Analysis

The balance sheet, which is also expressed as the table of the financial position of the company and is the starting point of the fundamental analysis, describes the resources (assets), liabilities and equities (net worth) of the companies. Unlike the profit & loss statement, it shows a snapshot of the company at a certain time. Namely, it is not related to the period spread over the annual process as in the profit & loss statement. Basically, it works in the context of the assets necessary for the company to continue its activities equal to the sum of the company's financial liabilities and equity. One of the main reasons why the balance sheet analysis is so important is that it shows whether the company has a balanced and stable structure enough to pay its current debt. Another critical reason is that analysis and liquidation of assets that can attract potential investors can be performed with this financial statement.

2.4.3 Cash Flow Statement Analysis

The cash flow statement is a financial statement that shows where a company gets its cash from and how it is spent. In its simplest terms, it is a summary of the amount of cash or cash equivalents entering and leaving the company. Essentially, cash flow consists of three key components: cash flow from operating activities, cash flow from investment and cash flow from financing. The cash that the company generates (or spends) as a result of its internal activities is analysed under the heading of operational cash flow. This section includes records of changes in net profit, working capital (e.g. changes in assets and liabilities). Cash flow for transactions on an investment basis includes financial items such as capital expenditures on equipment and property, sales of assets. Financial cash flow is dominated by debt and equity. It includes items such as repayments of short- and long-term debts and dividend payments.

2.4.4 Financial Ratio Analysis

Financial ratios, which can be used for many purposes, are one of the important elements in the process of evaluating the performance and financial condition of an entity. Financial ratios are numerical ratios obtained from three financial statements of companies. They are powerful tools that summarize the financial statements of companies, such as the company's ability to pay its debts and to possess managerial adequacy, and are used to measure the overall financial health of companies [7]. In this way, the ratios create a prediction about whether companies will remain "viable" in future. They allow monitoring of company performance from year to year and observing potential trends (changes). In addition, even though there are many financial ratio values, the ratios help to compare different companies' financial health more easily [8]. Thus, it provides a very effective solution for both internal and external users. Therefore, it is necessary to focus on certain ratios that can play a key role in measuring the financial health of companies. These ratios are liquidity ratios, leverage ratios, effectiveness ratios, profitability ratios, market value ratios, price multiples and valuation ratios.

Liquidity ratios: Liquidity ratios are financial measures that measure a company's ability to meet its short-term liabilities and cash flows without any capital increase. With this ratio, a company's ability to pay its debt obligations is measured by calculating the current ratio, acid-to-test ratio, cash ratio and operating cash flow ratio metrics.

Leverage Ratios: Like liquidity ratios, leverage ratios directly measure a company's financial health, but leverage ratios examine a company's ability to meet its long-term obligations. The main leverage ratios are as follows: debt ratio, debt-to-equity ratio and interest coverage ratio.

Efficiency Ratios: Efficiency ratios measure how effectively a company utilizes its assets. Thanks to these ratios, they can be used as an indicator of the efficiency of how much income is generated by using the assets. Common effectiveness ratios include: asset turnover ratio, inventory turnover ratio and days sales in inventory ratio.

Profitability Ratios: Profitability ratios are used to evaluate the company's ability to generate net profit according to its revenue, costs, assets and equities by looking at the data of the company over a period of time. In other words, profitability ratios are a measure of how effectively profits can be made and value produced for stakeholders.

Commonly used ratios are: gross margin ratio, operating margin ratio, return on assets ratio and return on equity ratio.

Market Value Ratios: Market value ratios are used to evaluate the current share price of a publicly held company's stock. This only helps potential investors understand whether the company's stock is overpriced or underpriced. The most commonly used market value ratios are as follows: book value per share ratio, dividend yield ratio, earnings per share ratio and price-earnings ratio.

Price Multiples: It gives the ratios of the current share prices of a publicly held company in terms of financial items such as earnings, sales and cash flow. Thus, it is possible for investors and analysts to make comparisons between companies and between different financial years of the same company. Commonly used ratios are: price to earnings ratio, price to sales ratio, price to book value and price to free cash flow.

Valuation Ratios: Valuation ratios are used in calculating the company's value. Commonly used ratios are: EV/EBITDA, EV/Sales, EV/FCF and book-to-market value.

2.5 Bankruptcy Analysis

The term bankruptcy is expressed as the inability of a company to pay its debts to its creditors [9]. The bankruptcy of a company and even the possibility of going bankrupt is important for the company's investors and society. Therefore, bankruptcy prediction should be made before the bankruptcy of a company and necessary and appropriate models should be built. In this part of the model, machine learning algorithms are used to predict whether companies will go bankrupt. In this way, it will be possible to predict the bankruptcy of companies with their financial statements and financial ratios.

2.5.1 Oversampling with SMOTE

Unbalanced data set is a problem that is frequently seen in classification problems and occurs when the class distributions are quite far from each other. This problem arises because the majority class dominates the minority class in machine learning algorithms. Owing to this, algorithms often predict the entire data set very poorly for the minority class, showing proximity to the majority class. Even though there are different metric selection and re-sampling methods to solve such problems, the SMOTE sampling method is the easiest and most useful method to apply.

SMOTE oversampling technique starts from the samples of the minority class and generates synthetic new observations in the feature space randomly by interpolation method. Thus, it balances the majority class with the number of observations. However, it does not interfere with the model other than increasing the number of samples and does not provide extra information to the model. According to some sources, the point to be considered while applying SMOTE is that the method should be applied only to the train data set and the original test data should be used while testing the data. Nevertheless, in some models, after all the data is balanced with the SMOTE method, the split of train and test data and the application of algorithms in this way also stand out in practice.

2.6 Solution and Analysis

The results about the financial fundamental analysis and ratio analysis of company *X* and its bankruptcy prediction model are given in this section.

2.7 Fundamental Analysis and Ratio Analysis

The profit & loss statement, balance sheet, cash flow statement and financial ratios of *X* company mentioned in 7 different categories in the model section will be given in this section.

First of all, the financial items of net sales, gross profit, total operating expenses, operating income, profit before tax, profit before change in accounting and net profit have been calculated in the profit & loss statement. In order not to adversely affect the general flow and visuality of the section, the profit & loss statement of *X* company is not given here. The relevant tables and code outputs can be examined in the Jupyter notebook of the chapter.

Company *X*'s balance sheet is also examined, and its total assets, total current liabilities, total non-current liabilities, total liabilities and total stakeholder equity items are calculated.

Finally, the cash flow statement of company *X* has been examined. Cash used for operating activities, cash used for investment activities and cash used for financial activities are calculated, and the total cash flow is found on a yearly basis.

2.7.1 Profit and Loss Statement Analysis

Profit & loss statement analysis was made, after the necessary cumulative calculations and adjustments in the profit & loss statement, balance sheet and cash flow statement of *X* company. Gross profit, total operating expenses and operating income are visualized with two different graphical displays. Similarly, profit before tax and net profit are compared on a yearly basis. Thus, the trend based on years is followed by analysing the profit & loss statement graphically.

Gross profit, total operating expenses and operating income are shown in Fig. 1. As it can be understood from this figure, it is seen that company *X* makes more sales from year to year and its operating expenses increase accordingly. However,

Fig. 1 Gross profit, total operating expenses and operating income with stacked chart

the noticeable increase in operating profit, especially after 2014, means that the company makes more sales in return for its expenses. In Jupyter notebook, these three financial items are more clearly described and the increase in operating income is more evident. In order not to adversely affect the general flow and visuality of the sections, a symbolic visual is presented in each of the profit & loss, balance sheet and cash flow statements of *X* company. Other related graphics can be examined through Jupyter notebook of the chapter.

Finally, the pre-tax profit and net profit of company *X* are given in Jupyter notebook. There is a significant exponential increase in the net profit of the company after 2014.

2.7.2 Balance Sheet Analysis

Secondly, the balance sheet of company *X* has been analysed. First of all, the distribution of total assets, total liabilities and total stakeholder equity has been examined. Then the distribution of all asset types is visualized. In the third chart, total liabilities and total stakeholder equity are plotted on a multiple axis. In the fourth chart, the distribution of current and non-current liabilities is given and the sub-distributions of both types of liabilities are drawn. The distribution of stakeholder equity types is found in the last chart. However, these graphics are not presented in the text section and should be followed through Jupyter notebook. Symbolically, the chart of assets, total liabilities and stakeholder equity are given here.

Figure 2 shows the graph of assets, liabilities and stakeholders' equity. The equation of the "Asset = Liability + Equity" equation can also be seen graphically. It

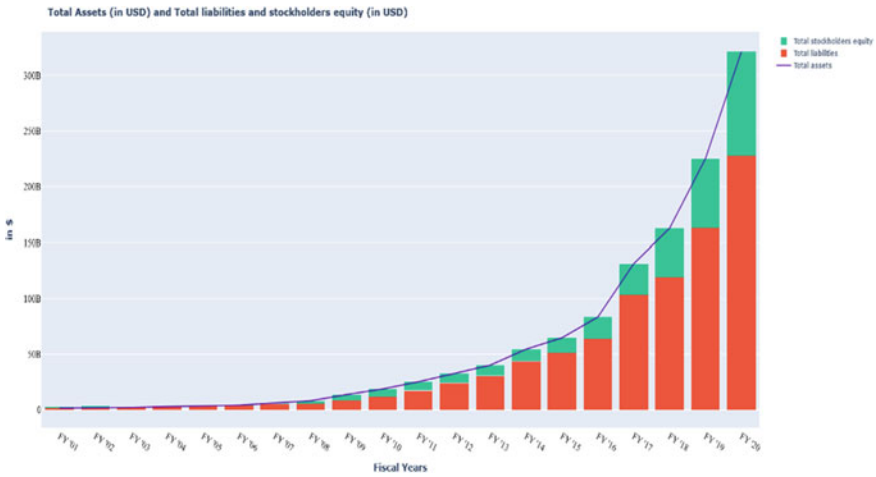

Fig. 2 Assets, total liabilities and stakeholder equity

seems that Company *X* has increased its assets tremendously over the years, and in turn, there has been a significant increase in its liabilities.

Afterwards, the distribution of asset items has been examined and shown. It is observed that property and equipment assets in particular have increased significantly from year to year. In addition, it is a remarkable case that the amount of market security assets can be easily converted into cash due to the speed of circulation in the market, especially in 2020. Cash and cash equivalents have also increased significantly from year to year.

In Jupyter notebook, the liabilities and stockholder equity of company *X* are shown on a 2-axis graph. While liabilities are followed on the left axis, equity capital is followed on the right axis. It seems that the stockholder equity tends to fluctuate relatively more over the years and increases more slowly than the liabilities. Generally, this situation is interpreted as a high risk for stakeholders. Yet, it would not be wrong to think that Company *X* increases its debts by following an aggressive policy in order to be more competitive.

Additionally, the distribution of current and non-current liabilities expressing long-term and short-term liabilities has been given, and the distribution for sub-financial items of both liabilities is also shown. It is realized that the long-term rental debt of *X* company and its debt to suppliers—as a short-term debt—(Payable accounts) have increased significantly over the years.

The equity distribution of company *X* is also shown in Jupyter notebook. The increase in retained earnings, which parallels the increase in net income, especially after 2015, is enormous. It can be deduced that Company X did not distribute dividends after 2015.

2.7.3 Cash Flow Statement Analysis

The distribution of operational, financial and investment cash, which is the source of total cash, has been examined, and the total cash flow has been plotted according to years.

In Fig. 3, it is seen that cash from operational activities gained a serious momentum after 2008. At the same time, it seems that company *X* increased its investments in the same year and its investment cash flow continued to grow negatively.

In Jupyter notebook, cumulative cash flows were shown. Except for the years 2001 and 2005, it is figured out that company *X* has a positive cash flow and the general trend from 2001 to 2020 is in the upward direction.

2.8 *Financial Ratio Analysis*

In this section, the financial ratios of *X* company in seven different categories are drawn.

Liquidity ratios can be seen in Fig. 4. Even though the general trend over the 20-year period is that Company *X* has generally worsened in terms of liquidation, it is clear that the company can continue to exist financially. However, the fact that the cash ratio is even below 0.5 most of the time creates the impression that Company *X* might have had problems in finding the cash it needs to pay its debts in some periods. In particular, the decrease in the current ratio over the years means that the current liabilities have increased against the current assets.

Figure 5 shows the ratios regarding the solvency of company *X*. Particularly, the serious fluctuations in the interest coverage rate can be shown as evidence that the

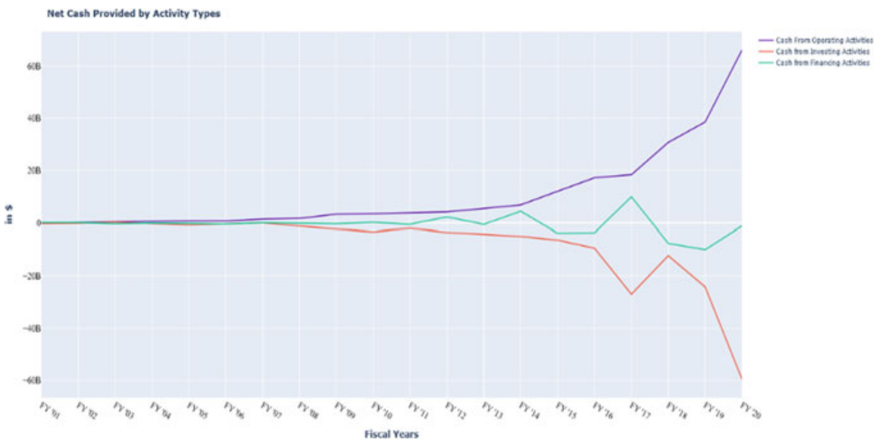

Fig. 3 Net cash provided by activity

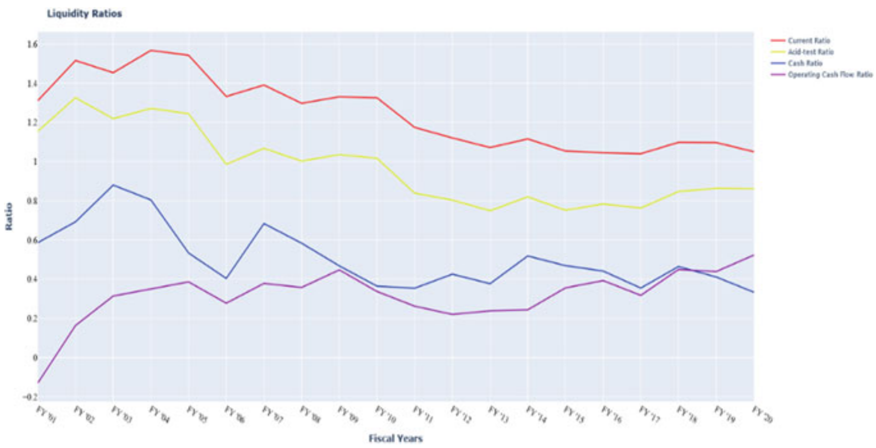

Fig. 4 Liquidity ratios

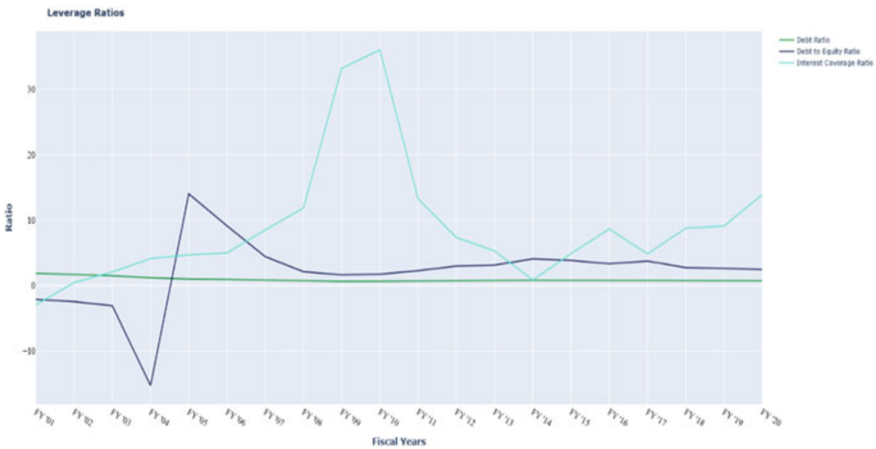

Fig. 5 Leverage ratios

company might have had difficulties in interest payments in some years. Also, a debt-to-asset ratio often above 0.7 means that most of Company *X*'s assets are financed by creditors.

The efficiency ratios are shown in Fig. 6. Both inventory and asset turnover ratios appear to be good overall. Assets and inventory appear to be handed over in relatively short times compared to the industry average. This is a piece of evidence of the efficiency of Company *X*.

In Fig. 7, profitability ratios are shown. Especially the good enough ROA and ROE ratios reveal the efficiency of the company's profit generating capacity.

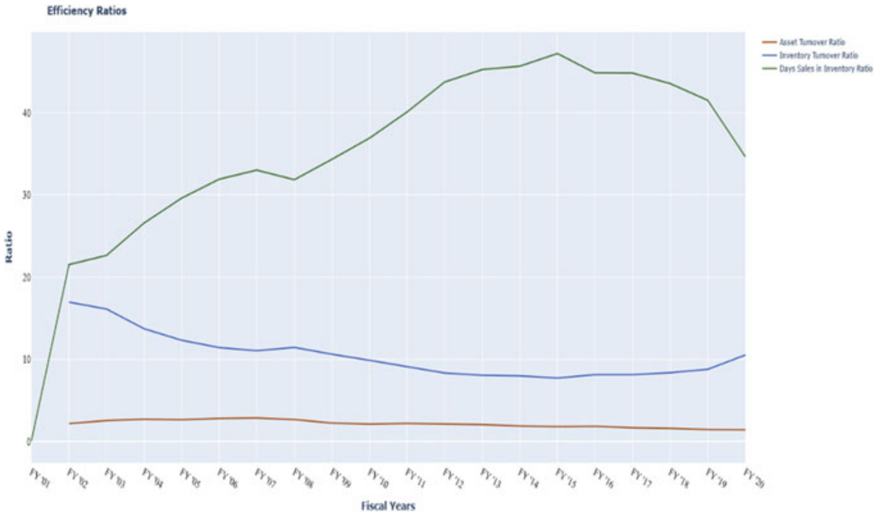

Fig. 6 Efficiency ratios

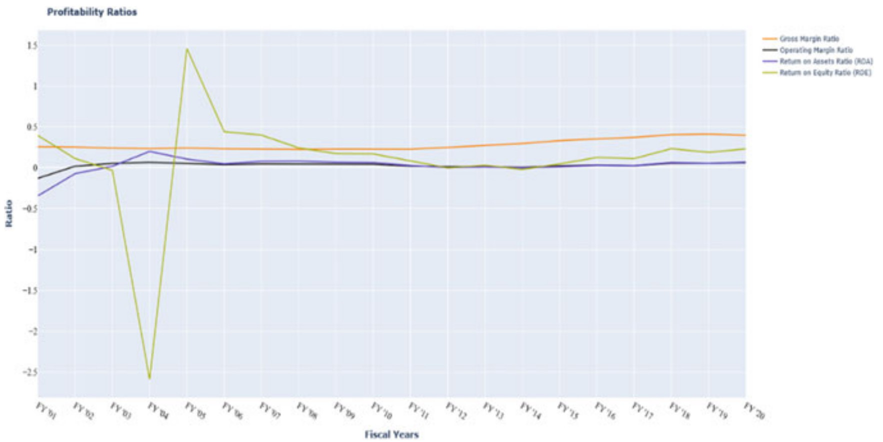

Fig. 7 Profitability ratios

In Fig. 8, there are ratios regarding the market value of Company X. Ratios such as earnings per share and sales show that the value per share of Company X has increased over the years. The significant increase in the book value, especially after 2015, is a proof of the increase in the net assets (total assets − total liabilities) of Company X per share.

Price multiples are given in Fig. 9. These ratios mainly concern investors who want to invest in company X.

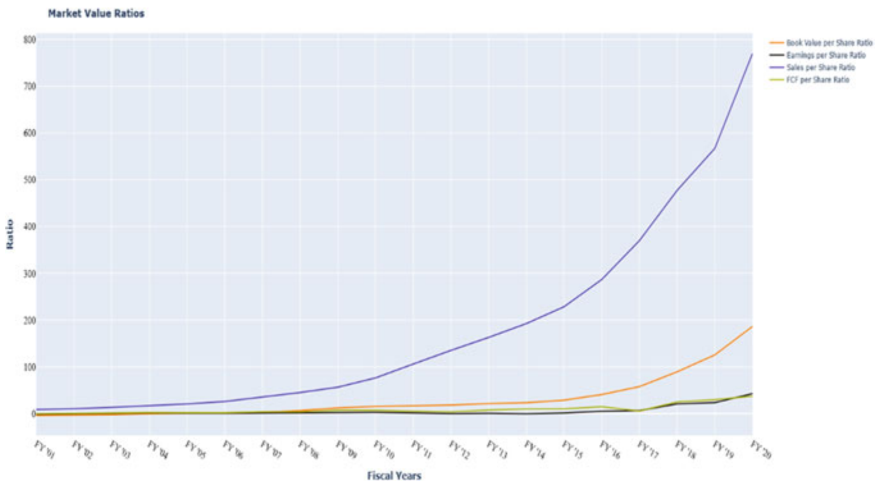

Fig. 8 Market value ratios

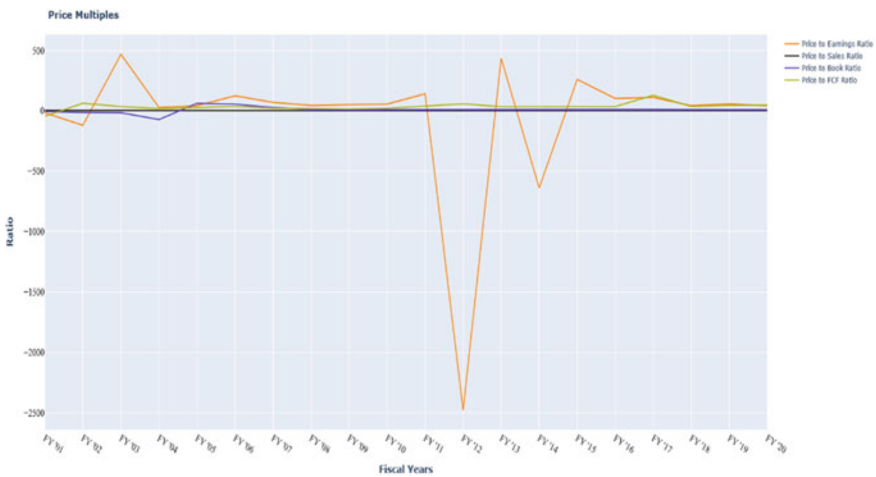

Fig. 9 Price multiples

Ratios presenting the overall value of the company are given in Fig. 10 under the name of valuation metrics.

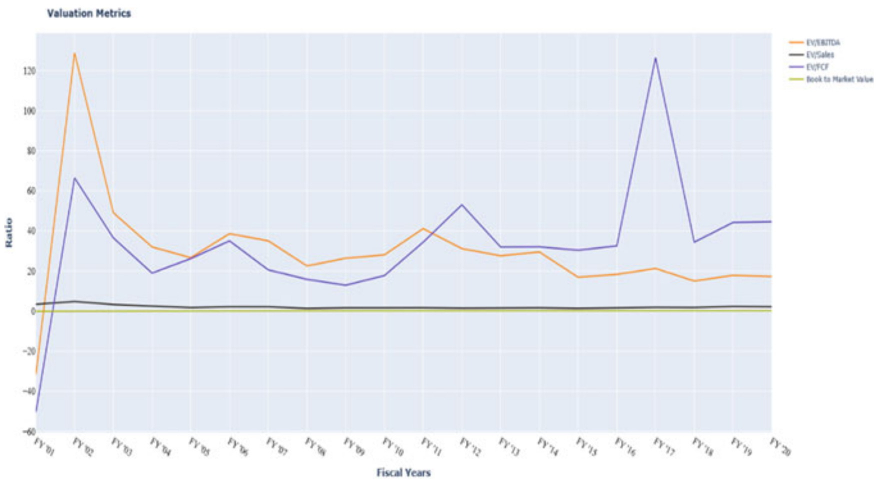

Fig. 10 Valuation metrics

2.9 Bankruptcy Prediction

In this part, the results of the bankruptcy prediction will be given in summary. First, the results of the pre-processed initial model will be shown. After the multi-collinearity control, the result of the data obtained after the SMOTE technique will be offered. These results are also divided into two and will be presented as the results of the original test data and the results of the test data applied to the SMOTE technique. After the dimensionality reduction, the results of the "reduced model" (with SMOTE used test data) will be given and compared with the original model. Finally, a prediction (bankrupt or healthy) will be made regarding the financial situation of Company *X* from 2001 to 2020.

First of all, the distribution of the data for 6819 companies in the data is shown in Fig. 11. As can be seen from the figure, there is a serious imbalance between classes.

In Fig. 12, the confusion matrix values of the initial models divided into the train–test data with 80–20 are shown. As can be seen from the matrices, none of the models can predict "class 1 data" well, since the amount of data with class 1 is very small. Therefore, it is clear that the data should be processed a little more and models should be developed with a different approach.

In the next stage of the analysis, multi-collinearity control was performed and 12 features with high correlation with each other were excluded from the analysis. Afterwards, the sample has been increased with the SMOTE technique and a balance has been achieved between the classes. This is shown in Fig. 13.

Then, all data have been trained with 12 models using the train data. The point to note in this regard is that SMOTE is applied only to the train data in the initial data. That is, the test data is kept as the original data at this stage. The sample has been expanded with SMOTE and then the data has been trained with a larger sample,

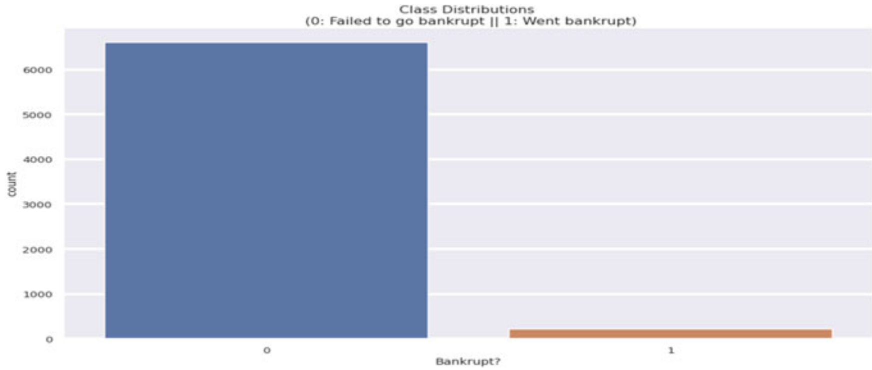

Fig. 11 Class distributions

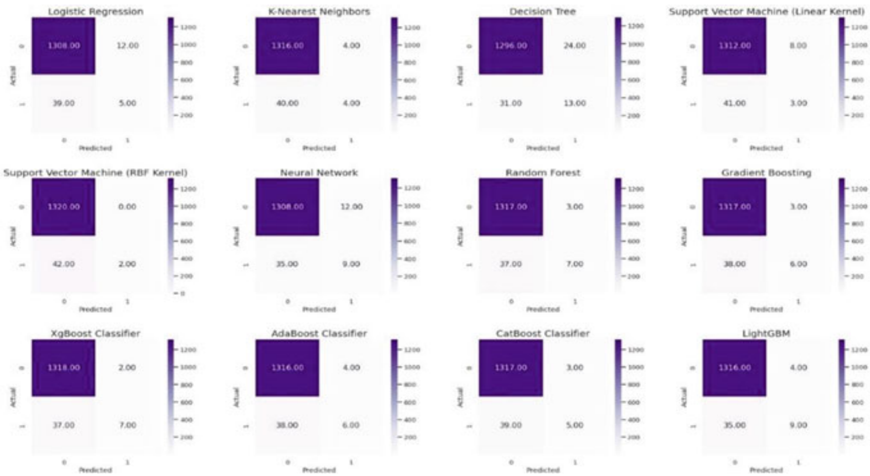

Fig. 12 Confusion matrix of initial models

but the models have been tested with the original test data. Since the models were trained with more data, the test data has been predicted to be better than the initial model. Figure 14 gives the ROC curve of the models, and Fig. 15 gives the confusion matrices. As can be seen from the ROC curve and confusion matrices, while logistic regression, SVM methods and AdaBoost classifier give relatively good results, the results of other models are only mediocre.

In the next model, the SMOTE technique has been applied directly to the raw data (before train–test data split). In this analysis, SMOTE technique has been applied to the raw data. Afterwards, train and test data have been created. Thus, the trained models were validated with the test data, which has a more balanced class distribution. Nevertheless, due to the loss of originality of the test data, there was a problem of overfitting in the models. Figure 16 shows the ROC curve of the model, and Fig. 17

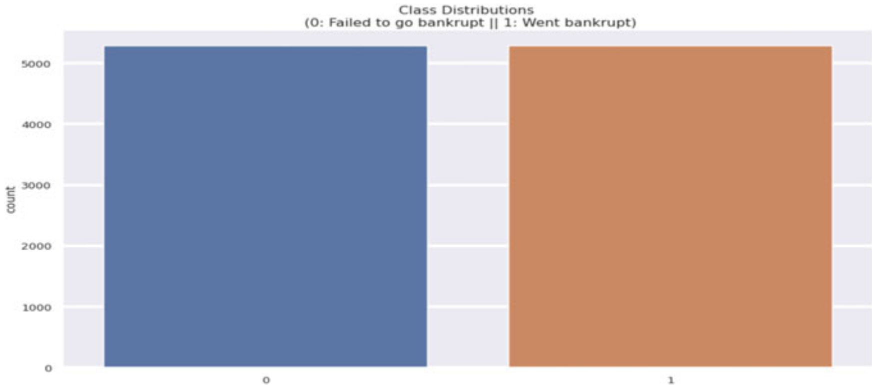

Fig. 13 Class distributions (original test data)

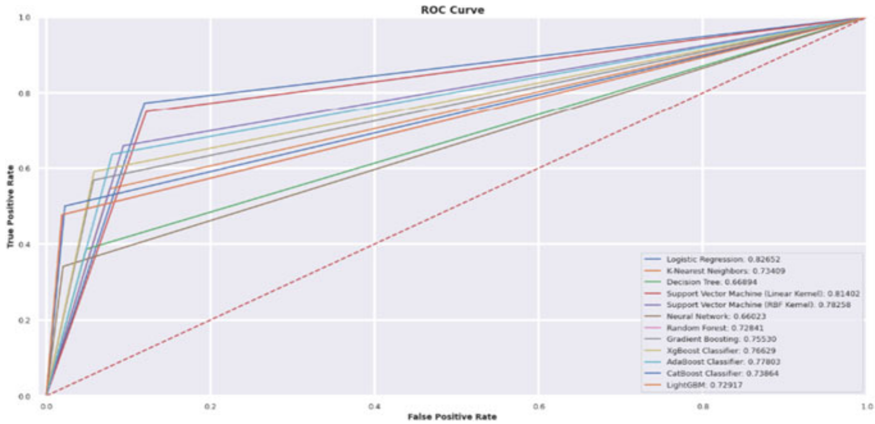

Fig. 14 SMOTE model ROC curve (original test data)

shows the confusion matrices. As can be seen from both performance metrics, many models predicted classes with around 95% success. Especially the success of models such as neural network, CatBoost, LightGBM is very high.

Afterwards, dimension reduction has been applied in order to express the model with fewer attributes. In this way, it is aimed to establish a decent enough model with fewer variables. As a result of the analysis depicted in Fig. 18, the contribution of the variables to the model has been shown and a "reduced model" has been established with 4 attributes according to the graph. Since the marginal contribution of the features after the 4 features to the model has decreased noticeably, it is enough to select 4 features.

After the dimension reduction was performed, the new data with 4 variables was rerun with the SMOTE model (in which the test data was also SMOTE). ROC curve and confusion matrices gave relatively poor results compared to the original model,

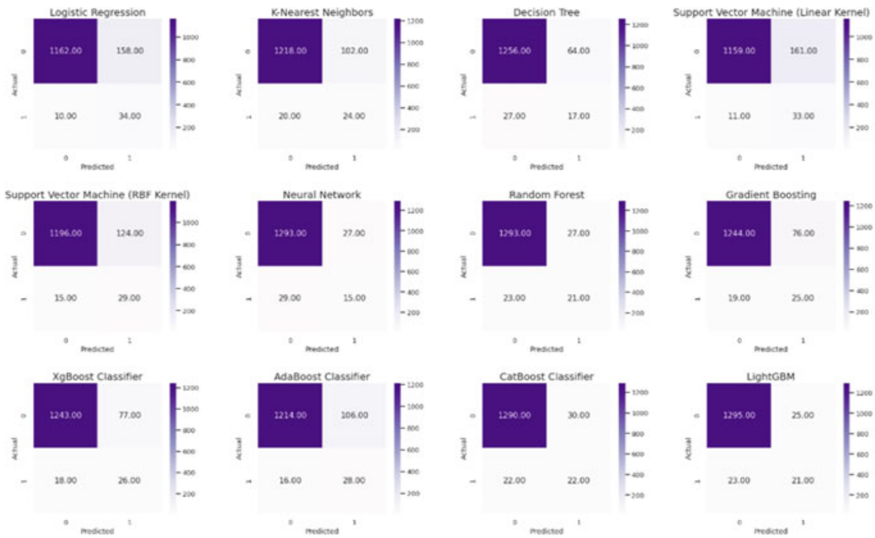

Fig. 15 SMOTE model confusion matrices (original test data)

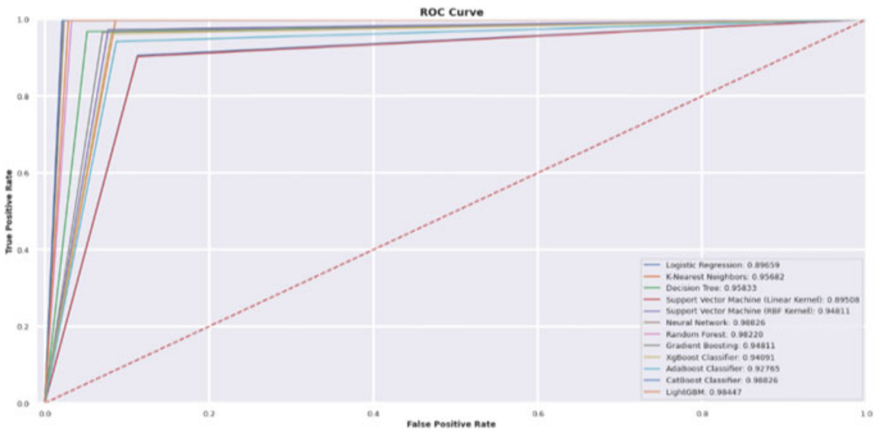

Fig. 16 ROC curve (SMOTE test data)

but despite the fact that the model could be expressed with only 4 variables instead of 62, no tremendous performance degradation was observed. Figure 19 shows how the performance of the 12 models changed after dimension reduction. In spite of the decrease of about 10% in some models, it can be said that it is quite satisfactory to establish models with only 4 features.

In the last part of the analysis, all features in the bankruptcy forecast data were calculated using the 20-year data of company *X*. This data of company *X* was given to both original models (original test data and SMOTE test data models) and the

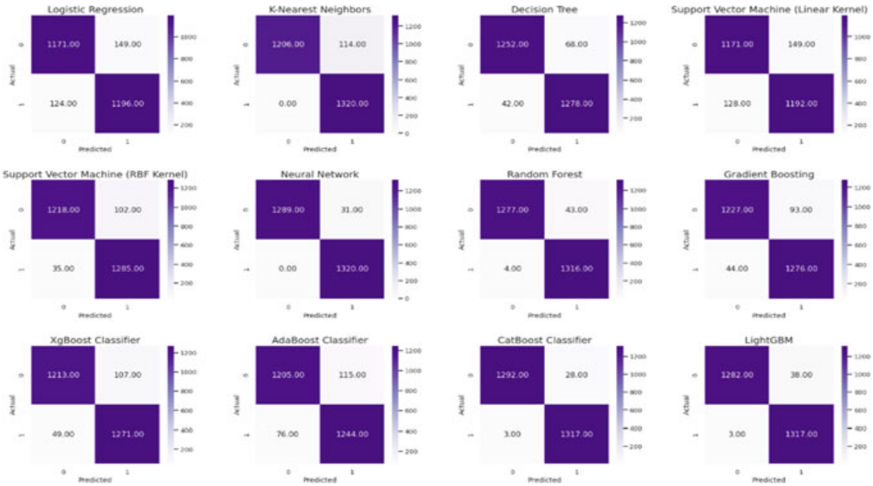

Fig. 17 Confusion matrices (SMOTE test data)

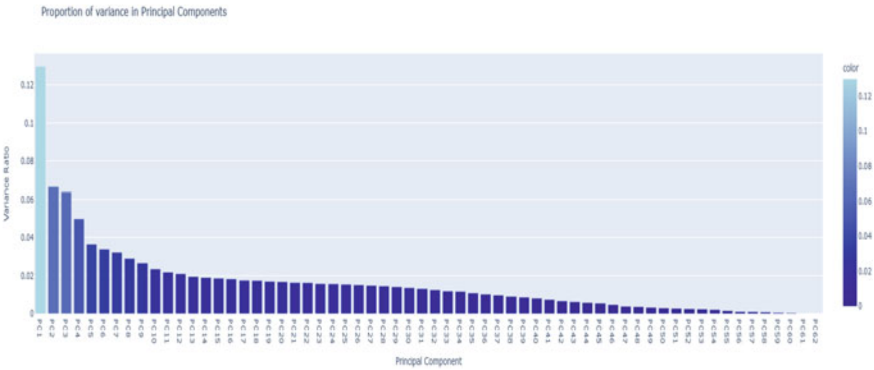

Fig. 18 Proportion of variance in principle components

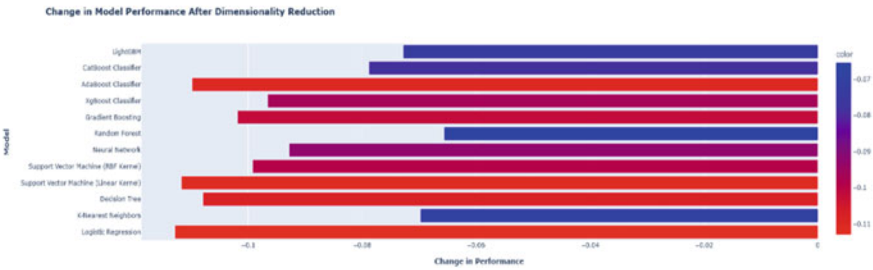

Fig. 19 Change in model performance after dimensionality reduction

prediction was made. Since the tables are difficult to read in the text, references are made to the Jupyter notebook. The results of the models using the original test data are also given there. The majority of algorithms in this model did not show Company X as insolvent for any year. However, a significant portion of the algorithms, again analysed using the SMOTE used test data showed that Company X went bankrupt in some years.

Finally, the bankruptcy prediction of company X was fulfilled by using the ensemble method which combines more than one model and generally gives better results than a single method. Ensemble was applied for both SMOTE and original test data. Accordingly, the financial situation of company X in the past years is shown in Jupyter notebook.

3 Conclusion

In this case study, first of all, comprehensive profit & loss statement, balance sheet and cash flow statement of X company are analysed. Afterwards, the necessary financial ratios have been calculated and the fundamental analysis has been completed. In the second stage of the case study, a bankruptcy prediction model using 12 machine learning algorithms has been developed as a result of data obtained from more than 6800 companies. This model is also supported by 3 sub-models created within itself. This case study can be developed by selecting the best parameters for the 12 machine learning algorithms and by doing cross-validation.

4 Credit Risk Analysis

4.1 Problem Definition

Credit risk analytics is a type of analytics or techniques that investigate and analyse the financial history of a person, institution or a company whether they can use and apply a credit for some purposes, or they are good at pay their credit back on time. Credit risk analytics can use different models to evaluate the applicant's financial situations. These models vary during analyses according to type of data and number. In the following case study, first the problem is going to be introduced and data will be prepared for analysing the applicants' situations and different models have been used to evaluate the customers' financial situations. After evaluation results, which model is the best model to be used in the credit risk analytics. In the final stage, a brief theory of the best model has been introduced.

4.2 Case Study

The German Bank wants to determine whether customers who apply for credit can be approved to give credit. The bank desires to know information about customers' economic conditions, such as how much their salaries are, for what customers want to take credits (i.e. buying a car, having a house, a tv or a mobile phone, etc.) and so on. If the applicant is in good condition to pay back money, the bank opens a credit to customers who apply the bank credit services. The case is how to make a faster decision that giving a credit to a new customer is risky or not instead of investigating one by one each applicant. According to previous applicants, that is the research question. All it is needed to find the most accurate model to evaluate the applicant credit approval. The German Bank System has applicants to banks to take credit. All data has been obtained from [10]. The credit approval services demand some information about applicants and keep this information to evaluate accepting or rejecting credit applications. Following information is about the applicants' features the German Bank System's requirements:

- Age: This feature is numerical information and legal age for applying the credit.
- Gender: This is string and categorical information that has two categories, namely male or female.
- Job: Job feature is a categorical feature that has four categories which are expressed numerically instead of categorical strings; 0: unskilled and non-resident, 1: unskilled and resident, 2: skilled, 3: high skilled.
- Housing: This feature is string variable which means categorical that these are three categories which are own, rent and free.
- Saving Account: This is a string variable that means categorical variable indicates if applicants have bank account, there are four categories; little, moderate, quite rich and rich.
- Checking Account: This is a categorical variable with three categorical class; little, moderate and rich.
- Credit Amount: This is a numeric variable that how much money applicants apply on Euro.
- Duration: This feature is a numeric variable to show how many months the applicants want to pay back credit.
- Purpose: This feature is a string variable that means categorical variables consist of eight categories. That is the aim of taking credit such as car, furniture/equipment, radio/TV, domestic appliances, repairs, education, business, vacation/others.
- Risk Situation: This feature is value target that string with two categorical variables as good or bad situation.

For the research question, original data sets include 1000 entries and 20 features. But in this case, only few of features are used to analyse to model research questions. Used features have been described in the problem definition section. There are ten features used in the analysis of this case which are age, gender, job, housing, saving account, checking account, credit amount, duration, purpose and risk situation.

The aim of this case study is to find out what algorithm or machine learning method gives the best accuracy or the best performance. The bank credit approving department uses the algorithm to make more fast and efficient method without losing time on approvals.

4.3 Data Information

In this section, first it is necessary to explore the data and its features. Data has ten features columns, and some of them are numerical variables and the rest of them categorical. Since categorical ones are sex, job, housing, saving accounts, checking account, credit amount, purpose and risk. numerical variables are age and duration. From these variables, risk categorical variable has been assigned to the dependent variable as binary classification good referred 1 that takes the credit and bad referred 0 that does not take the credit. This risk variable means that situation of a given customer is good or bad. Since customers can be approved giving credit defined as good which is accepted, otherwise defined as bad condition which is rejected. Apart from the risk feature, rest of the features have been assigned to independent variable (Table 1).

In the case study, there are seven methods that are applied to the data and get the results. These methods are decision tree classifier, random forest, naive Bayes, *k*th nearest neighbour, logistic regression, support vector machine and XGBoost classifier as machine learning methods. All these methods are called classification methods that means labelled as supervised learning methods on machine learning field.

There are missing values in data sets, there are two types of implementations of algorithms. One of these types is to clean missing data and another one is to proceed without cleaning missing values. These approaches are common in the machine

Table 1 Number of features non-null values

#	Column	Non-null count	Dytpe
0	Unnamed: 0	1000	Int64
1	Age	1000	Int64
2	Sex	1000	Object
3	Job	1000	Int64
4	Housing	1000	Object
5	Saving accounts	817	object
6	Checking account	606	Object
7	Credit amount	1000	Int64
8	Duration	1000	Int64
9	Purpose	1000	Object
10	Risk	1000	Object

learning field, and they have some advantages and disadvantages. One of the most popular advantages is all information for modelling is available that algorithms are easily run. One of the most well-known disadvantages is removing missing values from the data will cause loss of the information. Most researchers do not want to lose information to implement algorithms. In this case, all algorithms and methods have been applied with missing values and without missing values.

4.4 Data Cleaning and Preparing

First, all algorithms will be applied to cleaned data. Missing values have been cleaned with their rows. Saving account and checking account features have missing cells. Saving account feature has 183 missing cells, and the checking account has 394 missing cells. After cleaning missing values on both features, there are 522 rows left. Why 522 rows are that there are some missing values intercepted with saving accounts and checking accounts. All data has been cleaned but not prepared yet for modelling and evaluating. For evaluation, it is necessary to group data as a one binary target variable and the rest of the features are going to be independent variables. Risk feature has been selected as target variable that is a categorical variable as defined before. This risk categorical variable constituted from good and bad categories means that the good category means that credit application approved; on the other hand, bad category means that credit application not approved. This risk category defines binary classification that good is 1, bad is 0.

In the second stage, categorical independent variables have been group by numbers to make categories and assigned using dummy variables method. In order to prevent from multi-collinearity problem, first of columns has been deleted for each explanatory variable group. In the beginning data has 10 features that one of the features is assigned the target variable, which is risk, rest of them are explanatory variables which are age, credit amount, duration, sex, job, housing, saving accounts, checking accounts and purposes. There are nine explanatory variables are separated by three of them are numerical variable six of them categorical variables. To implement models to the data, categorical variables must be grouped by dummy variables that make categorical variables to numerical variables. After organizing data with dummy variables, 27 features have been gotten. Normally, nine features have been in data but with dummies with one-columns deleting totally of 21 explanatory variables have been displayed. Once cleaning and preparing data following structures have been founded for target variables and independent variables. These independent variables features are as given in Fig. 21 (Table 2).

The above explanatory variables are after the cleaning and preparing the data. There are 522 entries and explanatory variables. Finally, the data has been ready to be run.

There are a lot of methods to make preparation data to be modelled. One of them has been already mentioned above is to remove missing values. Since removing data is going to cause loss of information, many researchers do not desire to apply this

Table 2 After dropping null cells

#	Column	Non-null count	Dytpe
0	Age	522	Int64
1	Credit amount	522	Int64
2	Duration	522	Int64
3	Sex_female	522	Uint8
4	Job_1	522	Uint8
5	Job_2	522	Uint8
6	Job_3	522	Uint8
7	Housing_own	522	Uint8
8	Housing_rent	522	Uint8
9	Saving accounts_moderate	522	Uint8
10	Saving accounts_quite rich	522	Uint8
11	Saving accounts_rich	522	Uint8
12	Checking account_moderate	522	Uint8
13	Checking account_rich	522	Uint8
14	Purpose_car	522	Uint8
15	Purpose_domestic appliances	522	Uint8
16	Purpose_education	522	Uint8
17	Purpose_furniture/equipment	522	Uint8
18	Purpose_radio/TV	522	Uint8
19	Purpose_repairs	522	Uint8
20	Purpose_vacation/others	522	Uint8

method to data sets. Because, in order to clean missing values, one needs to remove all columns or rows that this approach is going to corrupt lots of other information including other features. Instead of removing the missing values, all missing values can be filled with one of some sensible approaches. This without dropping missing values analysis separated into two sub-methods. One of the methods is to run models that fill empty cells without knowing which model fill cells, the other one is to first fill cells according to the most sensible method in the literature. In the literature, there are many filling techniques that help to avoid data loss [11] which is called imputation. Imputation means filling missing values with some numbers or some categorical properties.

According to the type of data, there are some preferred imputation techniques that can be used to prepare data to be modelling. Data has two categorical variables which are saving account and checking account. To fill missing values cells, maximum repeating sub-feature impute the empty cells. Saving accounts missing cells are filled with the little categorical values because most of the customer approximately more fifty per cent have little accounts. Checking account missing cells are filled with the moderate categorical variable that checking account customer mostly have

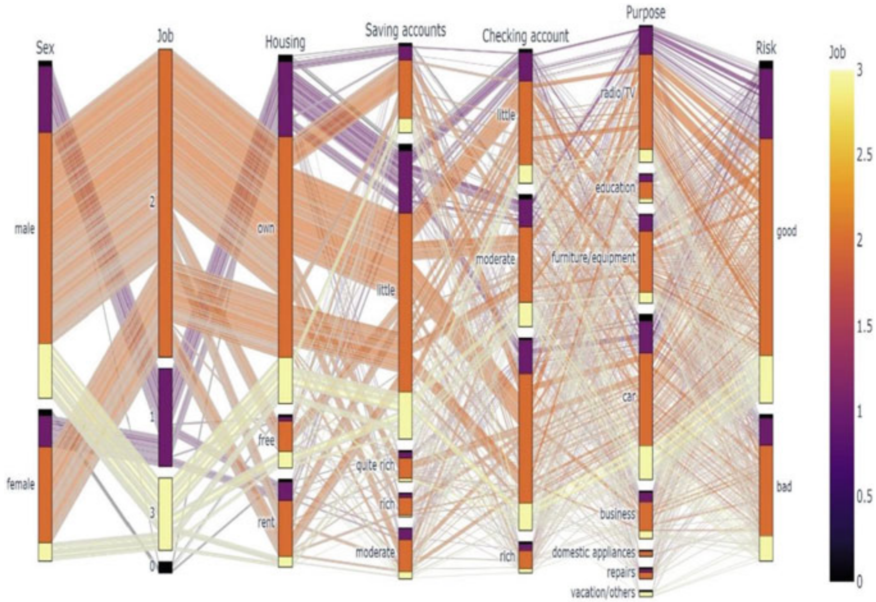

Fig. 20 Parallel diagram of the categorical features

moderate account more than eighty per cent of the applicants. With imputation of the missing values, data have been prepared for the modelling.

According to parallel categorical diagram, most of the applicant's gender is male compared to female. From the diagram, generally people work second job type and mostly have their own house. The applicants tend to have quite little saving accounts in the bank and little and moderate checking accounts. They apply mostly to buy TV/radio, furniture/equipment, car. Risk situations are good even if some of them are bad situation. All information can be gotten from Fig. 20.

4.5 Model

After the data have been cleaned and prepared for the modelling, seven machine learning algorithms have been applied which are naive Bayes (NB), decision tree (DT), random forest (RF), logistic regression (LR), K-nearest neighbour (KNN), support vector machine (SVM) and XGBoost (XGB) classifier models have been run. The aim of the applying seven classification algorithm is to find out what algorithm gives the more accurate result. There are three data types that models have been implemented; first algorithms are applied with dropping missing values, and then second all algorithms are applied without dropping any of the missing values. And lastly, all algorithms are run with imputed cell values.

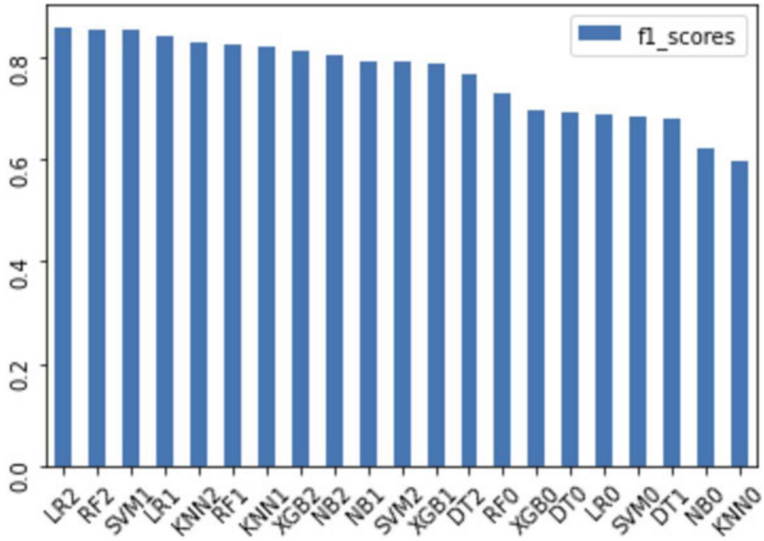

Fig. 21 F1 scores bar chart for all situation on models

For the evaluation of the model performance, classification matrix has been used by a matrix. In the model evaluation, F1 score model evaluation metric has been considered because of the data that give more accurate results for the models. From the model performance analysis, after the missing value imputation applied to data, logistic regression has given the best performance metric among the classification methods that have been already mentioned which these models are LR, DT, RF, NB, SVM, KNN and XGB models.

All measures of F1 scores have been compared in the following table (f1_scores are the approximate values).

It can be understood that from Fig. 21, LR2 model has the highest f1_score among the rest of the other applied models.

As it can be seen from Table 3, the highest f1_score is found from the LR model overall when it is tested with different test and train data sets cover 80 per cent of data and 20 per cent of data is test data with random state is 7. As a result, the LR model has been selected by the German Bank to apply which decision has made new customer applications. There are other models which also have enough f1_score to use but LR is the highest one even if there is no big difference.

Table 3 F1 scores for all situation on models

F1_score	Dropped missing values	Without dropped missing values	Imputation with maximum repeating categorical values
NB	0.622	0.792	0.806
DT	0.692	0.679	0.767
RF	0.730	0.824	0.855
LR	0.690	0.840	0.859
KNN	0.598	0.819	0.827
SVM	0.682	0.854	0.790
XGB	0.696	0.788	0.811

5 Investment Analytics

5.1 Introduction to Portfolio Analytics

One of the most fundamental problems of businesses today is to provide an efficient financial management. One of the main issues that businesses should be interested in the field of finance is the creation and management of an investment portfolio. This issue has become even more important in a complex environment such as a globalizing society, rapidly increasing competition and extensive economic changes at the national and international levels [12].

In 1952, Markowitz (Markowitz) developed the mean variance model that formulates the mathematical relationship between the returns and risks of assets [13]. In the following years, portfolio creation methodologies have diversified, with other models trying to expand it and eliminate its weaknesses. Thanks to these models, investors began to create portfolios that support certain investment styles and preferences.

In this section, besides the basic financial calculations, performance metrics and portfolio construction methods will be presented. In addition, the portfolios created with different models for the five technology stocks traded in the S&P 500 index will be compared with the performance metrics whose definitions are given.

5.2 Return & Risk Calculations

Return can be defined as how much gain or loss at time $t + 1$ from time t. It can be formulated as follows:

$$\text{Return}_t = \frac{\text{Price}_t - \text{Price}_{t-1}}{\text{Price}_{t-1}} \tag{1}$$

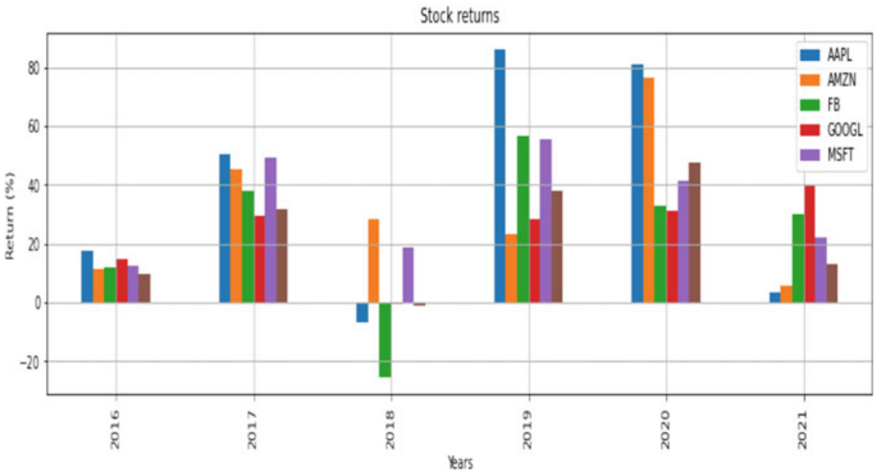

Fig. 22 Yearly stock returns

The cumulative return is an aggregate return or the total change in the investment price over a period of time. Along with the cumulative return, the compounded return is the rate of return that takes into consideration the compounding effect of the investment for each period (Fig. 22).

The risk of an asset is often expressed as its volatility, and it is usually defined as the standard deviation (or variance) of the returns on the asset (Table 4).

Similarly, the volatility of a portfolio can be calculated by the variance, weights and correlations of the assets it contains. Portfolio variance is formulated as follows:

$$\sigma_p^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_{1,2} \tag{2}$$

σ_p^2: Variance of the portfolio.
w_n: Weight of the nth asset.
σ_n^2: Variance of the nth asset.
$\sigma_{1,2}$: Covariance between assets 1 and 2.

Table 4 Expected returns and volatility of the stocks

Stock	Expected return (%)	Volatility (%)
AAPL	36.98	30.20
AMZN	37.28	30.32
FB	25.16	32.95
GOOGL	20.24	26.34
MSFT	31.92	27.76

5.3 Performance Metrics

Sharpe Ratio: It is the most widely used measure of risk-adjusted rate of return. It is calculated by subtracting the portfolio return from the risk-free asset's return and dividing it by the portfolio's standard deviation.

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p} \tag{3}$$

R_p: Return of the portfolio.
R_f: Risk-free rate.
σ_p: Standard deviation of the portfolio.

Sortino Ratio: It is the modified version of the Sharpe ratio but uses a different standard deviation, downside deviation.

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_d} \tag{4}$$

σ_d: Standard deviation of the downside.

Treynor Ratio: It is also known as the reward ratio for risk taken. The risk expressed in the Treynor ratio refers to the systematic risk measured by the beta of a portfolio, which will be explained in the next section.

$$\text{Treynor Ratio} = \frac{R_p - R_f}{\beta_p} \tag{5}$$

β_p: Beta of the portfolio.

Drawdown: It is defined as percentage loss from the last maximum peak. Maximum drawdown can also be defined as the worst percentage loss from a market peak to the very lowest point. It is a strong indicator that measures the downside risk of the portfolio.

5.4 Factor Models

Capital Asset Pricing Model (CAPM): The model was developed by Sharpe [14], Linter [15] and Mossin [16] to explain why different securities have different expected returns. CAPM explains that every investment carries two different risks, systematic and unsystematic.

The risk called systematic risk or beta is the risk of being in the market and expressed as follows:

$$R_i = R_f + B_i\left(R_m - R_f\right) \tag{6}$$

Table 5 Beta of the stocks

Stock	Beta
AAPL	1.14
AMZN	1.04
FB	1.09
GOOGL	1.00
MSFT	1.13

R_i: Expected return of the asset i.
R_f: Risk-free rate.
R_m: Expected return of the market.
β_i: Beta of the asset i (Table 5).

In the CAPM, beta is used to describe the relationship between systematic risk and an asset's expected return. Higher beta indicates that the asset is more volatile than the market. The CAPM is also graphically illustrated as the security market line (SML). As a straight line intersecting the vertical axis at the risk-free rate, SML depicts the risk–return trade-off (Fig. 23).

Fama French Multi-Factor Model: This model is usually expressed as a three-factor model that improves the CAPM (single factor model). These three factors are market, size and value factors [17]. In the formula below, $R_m - R_f$ represents the market factor, SMB represents the size factor and HML represents the value factor.

$$R_i = R_f + \beta_m \left(R_m - R_f\right) + B_s SMB + B_h HML \tag{7}$$

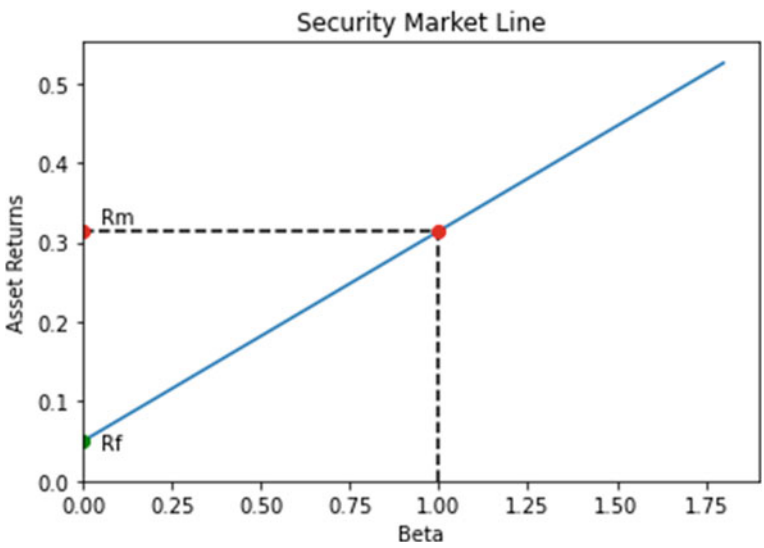

Fig. 23 Security market line of NASDAQ

R_i: Expected return of the asset i.
R_f: Risk-free rate.
R_m: Expected return of the market.
β_m: Market beta of the asset i.
β_s: Size beta of the asset i.
β_h: Value beta of the asset i.
SMB: is the portfolio returns of small stocks minus the portfolio returns of big stocks.
HML: is the portfolio returns for high book-to-market value minus returns of low book-to-market value stock.

5.5 Modern Portfolio Theory

Portfolio selection includes mathematical models designed to best meet the needs of the investor and investors seek the highest possible return against minimum risk. Therefore, they want to maximize risk-adjusted returns. Markowitz [13] firstly introduced the modern portfolio theory (MPT) that helps to determine the minimum risk level that the investor should undertake in order to reach the targeted return level. The Markowitz's MPT is often generically known as the mean–variance optimization framework. In this section, MPT will be used to create optimal portfolios that meet the needs of investors [18].

The efficient frontier is a graph showing the rate of return corresponding to volatility. It includes optimal sets of portfolios that offer the highest expected return for a given level of risk or the lowest risk for a given expected return level. If the points are derived by assuming all possible weights with different combinations of entities, the efficient frontier chart is arrived (Fig. 24).

If a risk-free security is available, then an investor's portfolio optimization problem can be formulated as follows.

$$\text{minimize}\frac{1}{2}w'\sum w \tag{8}$$

$$\text{subject to}\left(1-\sum_{i=1}^{n}w_i\right)r_f+w'\mu \tag{9}$$

w: is the vector of portfolio weights.
$\sum$: is the covariance matrix of the asset returns.
r_f: is the risk-free rate.
μ: is the expected returns (Fig. 25).

Minimum volatility portfolio aims to minimize the volatility of the portfolio, and it ignores expected return and focuses only on the risk. It is formulated as follows.

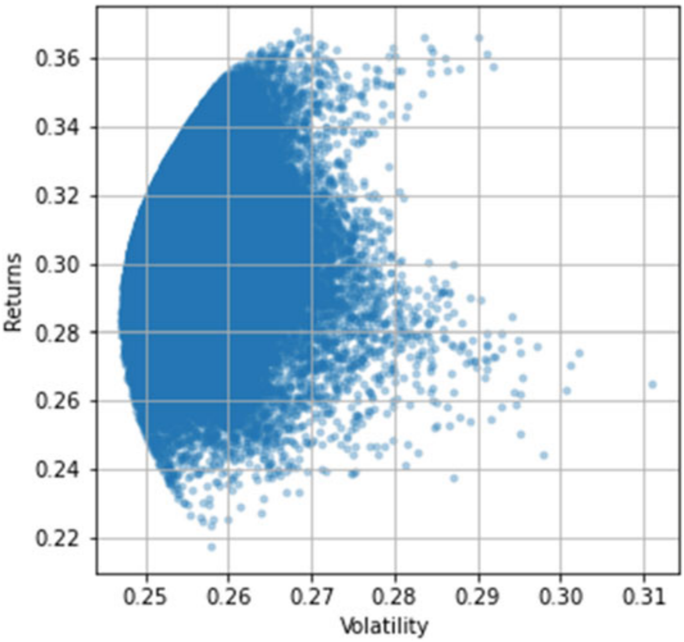

Fig. 24 Efficient frontier chart

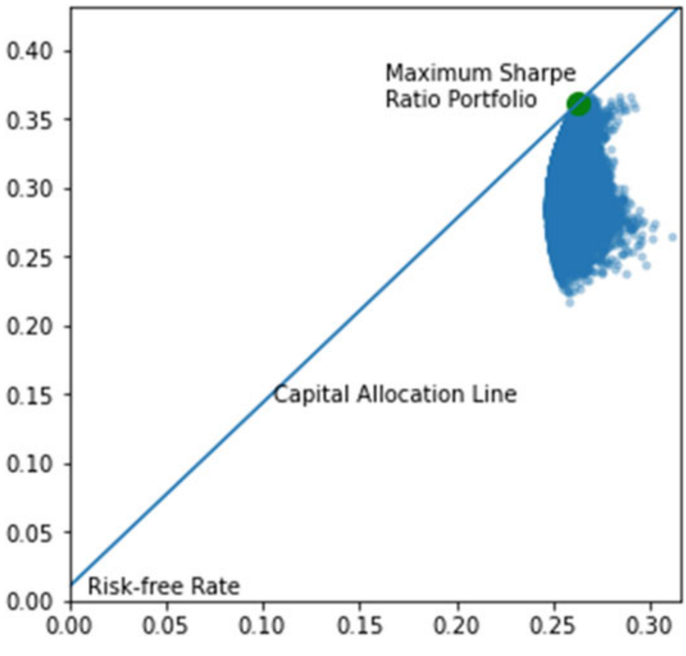

Fig. 25 Capital allocation line

Table 6 Results of portfolio optimization

Objective	Expected return (%)	Volatility (%)	Sharpe ratio	Sortino ratio	Beta	Treynor ratio	Max drawdown (%)
Equal weighted	30.29	25.16	1.16	1.43	1.08	0.27	77.77
Minimum volatility	28.31	24.71	1.11	1.36	1.06	0.26	75.20
Maximize sharpe	36.44	26.45	1.34	1.72	1.09	0.32	84.42
Efficient risk (25%)	31.75	25.0	1.23	1.53	1.08	0.29	79.60
Efficient return (35%)	35.00	25.78	1.32	1.66	1.1	0.31	83.02

$$\text{minimize } w^T \sum w \tag{10}$$

$$\text{subject to } 1^T w = 1 \tag{11}$$

Also, return can be maximized for a given target risk β_p. It is formulated as follows (Table 6).

$$\text{maximize } w^T \mu \tag{12}$$

$$\text{subject to } w^T \sum w \leq \beta_p \tag{13}$$

5.6 *Black and Litterman*

The Black–Litterman (BL) model is a portfolio allocation model that creates portfolios based on an investor's unique insights [19]. This model is also on the basis of mean–variance. In the model, expected returns are calculated as the new expected returns, which are found by the following expression.

$$E(R) = \left[\left(\tau \sum \right)^{-1} + P^T \Omega^{-1} P \right]^{-1} \left[\left(\tau \sum \right)^{-1} \Pi + P^T \Omega^{-1} Q \right] \tag{14}$$

$E(R)$: is the new expected return.
τ: is a scalar constant.
$\sum$: is the covariance matrix of the asset returns.

P: is a matrix that identifies assets involved in the views.
Ω: is the uncertainty matrix of views.
Π: is the prior expected return.
Prior expected returns in the above formula are calculated as follows.

$$\Pi = \lambda \sum w_{mkt} \tag{15}$$

λ: is the risk aversion coefficient or market-implied risk premium.
w_{mkt}: is the market capitalization weight of the assets.
Risk aversion coefficient is found by the following formula:

$$\lambda = \frac{R_m - R_f}{\sigma_m^2} \tag{16}$$

σ_m^2: is the variance of the market.
In the BL model, investors can provide different type of views. They can state "AMZN will return 30%" or "FB will drop 10%" as an absolute view. On the other hand, relative views can be stated like "AAPL will outperform GOOGL by 2%". In addition, investors can provide confidence level in their views.

In our case, absolute views will be provided as an example. Also, short sells will be allowed. Below table shows our views and also weights produced by the model using these views (Tables 7 and 8).

Table 7 Investor views and resulted weights in BL model

Stock	View (%)	BL weight (%)
AAPL	+20	55
AMZN	+30	107
FB	−10	−78
GOOGL	5	−20
MSFT	15	36

Table 8 Result table of BL optimization model

Objective	Expected return	Volatility	Sharpe ratio	Sortino ratio	Beta	Treynor ratio	Max drawdown
BL market risk aversion	48.02%	36.67%	1.28	1.88	1.09	0.43	91.61%

6 Financial Hedging Analysis

6.1 Problem Definition

Corporate risk management refers to all methods that companies use to reduce the risk they may face. Hedging is one of the financial risk management strategies that is defined as taking an offsetting position in derivatives market of related security. There are several ways to hedge current positions by using derivatives such as futures, options and forwards.

Futures contracts are agreements that secure a specified price at a specified time in future for two parties: seller and buyer. In this sense, futures contracts are commonly used for hedging currencies, stock market indexes and commodities. These contracts are traded on exchanges which regulate clearance and pricing of contracts. The price of future contract, which will be referred to as "futures", is the same regardless of which broker is used due to the centralized price mechanism.

Exporting and importing is a way to expand businesses with its risks such as foreign exchange risk. Internationally trading companies are usually exposed to foreign exchange risk, especially if the local currency is unstable and aims to limit the losses they may incur due to currency fluctuation. Companies aim to minimize this risk by taking a position in currency derivative markets that is called hedging. Financial risk manager of a company tries to find out answers for the following questions among available contracts [20]:

- Which futures contract is appropriate to use?
- When to take buy (long) or sell (short) position?
- What is the optimal hedge ratio for the amount exposed to risk?

Risk managers can determine the most appropriate answers to these questions by using hedging strategies. The aim of hedging is to neutralize the risk of exchange rate fluctuations and guarantee a fixed price for a specified future date. Futures are leveraged financial instruments in which investors only put in a margin which is a fraction of total notional value on contract. Futures have standardized specifications which are determined by exchange. These specifications are: (1) the asset: this is underlying product of future contract that can be a foreign currency or a physical commodity like gold; (2) the contract size: this is the amount of the asset that needs to be delivered within one contract; (3) delivery arrangements: the method or place of delivery for underlying asset that is specified by exchange; (4) delivery months: this refers to the month that the delivery occur; 5) price quotes: this refers that how price of contract will be quoted (6) tick size: this is the minimum price movement exchange allows; (7) position limits: this is the limits of daily price movements specified by exchange. As an example, Chicago Mercantile Exchange (CME) currency futures contract specifications are illustrated in Table 9.

Businesses or individuals open accounts with brokers to trade on futures and brokers direct their orders to exchange. This account is generally called margin account, traders are required to deposit an initial amount of cash specified by

Table 9 Contract specifications for foreign currency futures: Eurodollar [21]

EURO FX futures—contract specs	
Contract unit	125,000 Euro
Price quotation	US dollars and cents per Euro increment
Trading hours	CME Globex
	Sunday–Friday 5:00 p.m.–4:00 p.m. (6:00 p.m.–5:00 p.m. ET) with a 60-min break each day beginning at 4:00 p.m. (5:00 p.m. ET)
	CME ClearPort
	Sunday 5:00 p.m.–Friday 5:45 p.m. CT with no reporting Monday–Thursday from 5:45 p.m.–6:00 p.m. CT
Minimum price fluctuation	CME Globex
	0.00005 per Euro increment = \$6.25
	Consecutive month spreads: 0.00001 per Euro increment = \$1.25
	All other spreads: 0.00002 per Euro increment = \$2.50
	CME ClearPort
	0.00001 per Euro increment = \$1.25
Product code	CME Globex: 6E
	CME ClearPort: EC
	Clearing: EC
Listed contracts	Quarterly contracts (Mar, Jun, Sep, Dec) listed for 20 consecutive quarters and serial contracts listed for 3 months
Settlement method	Deliverable
Termination of trading	Trading terminates at 9:16 a.m. CT, 2 business day prior to the third Wednesday of the contract month
Settlement procedures	Physical delivery
	EUR/USD futures settlement procedures
Position limits	*CME position limits*
Exchange rulebook	*CME 261*
Block minimum	*Block minimum thresholds*
Price limit or circuit	*Price limits*
Vendor codes	*Quote vendor symbols listing*

exchange. They are also required to keep a maintenance margin which is the minimum amount of equity investors must hold in margin account. Gains and losses of futures are calculated every trading day with "marking to market" procedure.

An example of currency futures is given in the following case study.

6.2 Case Study

On 1 July 2021, ABC company from USA sells machinery to a German company and will receive €3.75 m on 25 September 2021. They expect that Euro will weaken

Table 10 Contract details of September Euro futures

Specification	Explanation
Contract size	125,000 Euro
Price quotation	US dollars and cents per Euro increment
Settlement process	Mark to market
Termination of trading	2 business day prior to the last Wednesday of the contract month
Delivery month	September
Initial margin	\$3000
Maintenance margin	\$2200

against US Dollar; therefore, they choose to take a position in the currency futures market to hedge their position. The current spot rate in the market is \$1.1764/1€ and futures contracts at \$1.170/1€ are available for delivery month. Contract details of Euro Futures are given in Table 10.

6.2.1 Setting up the Hedging

Under given conditions, the company sets up the hedging strategy against losses. The will be received in the end of September, and therefore, a future contract with same month is obtained. After the decision of expiry month, the appropriate position is taken since company is in long equity cash position, so they take put position in futures market. Lastly, company will decide the fraction of equity to be hedged. Investors can do either partial hedge or full hedge their current positions. The fraction of total equity hedged is called hedge ratio. In our case, different scenarios of hedge ratio will be applied to see gain/losses at expiry. If company aim to do full hedging, they need to put 30 futures contracts.[1]

The company can take position under these setups.

6.2.2 Simulating Spot Price

The company takes position in the currency futures market, and the settlement process is marked to market. The aim of this process is to enable investors to see their current financial situation based on current futures price. Therefore, the margin account will be updated based on the current situation. A futures price for Euro/USD futures is simulated and given in Fig. 26.

Based on the generated price, total gain/losses are marked to the margin account daily and first 10 days and last 10 days of contract holding period are given in Table 11.

Here in Table 11, there are 30 contracts and \$3000 is put in as initial margin for each contract, so \$90,000 in total. Margin account starts with \$90,000 and changes daily

[1] Number of contract for full hedging is calculate as: 3,750,000/125,000 = 30.

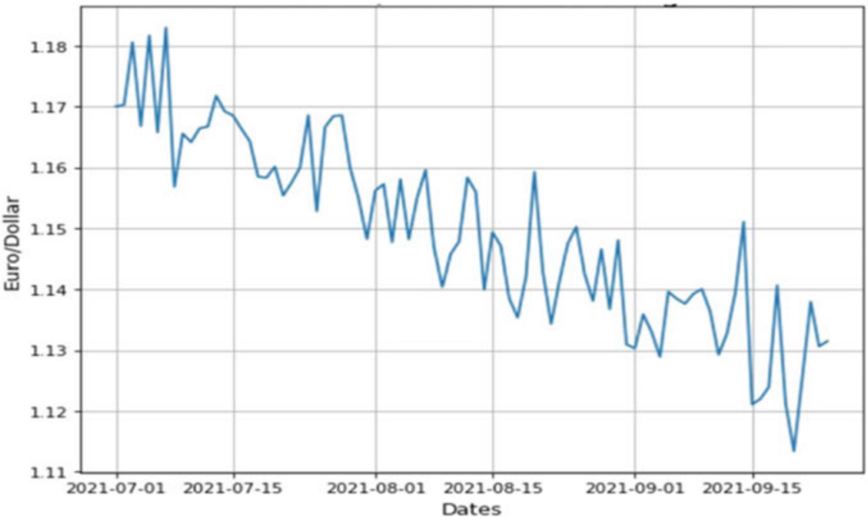

Fig. 26 Simulated euro/US dollar futures price

Table 11 Margin account balance

Dates	Futures price	Change in value	Gain/Loss ($)	Account balance ($)
2021–07–01	1.170000	–	–	90,000
2021–07–02	1.170292	0.0003	−1125	88,875
2021–07–03	1.180560	0.0103	−38,625	50,250
2021–07–04	1.166781	−0.0138	51,750	102,000
2021–07–05	1.181667	0.0149	−55,875	46,125
2021–07–06	1.165743	−0.0159	59,625	105,750
2021–07–07	1.182957	0.0172	−64,500	41,250
2021–07–08	1.156793	−0.0262	98,250	139,500
2021–07–09	1.165531	0.0087	−32,625	106,875
2021–07–10	1.164129	−0.0014	5250	112,125
⋮	⋮	⋮	⋮	⋮
2021–09–15	1.121084	−0.0299	112,125	274,125
2021–09–16	1.121991	0.0009	−3375	270,750
2021–09–17	1.123935	0.0019	−7125	263,625
2021–09–18	1.140673	0.0167	−62,625	201,000
2021–09–19	1.121300	−0.0194	72,750	273,750
2021–09–20	1.113382	−0.0079	29,625	303,375
2021–09–21	1.125451	0.0121	−45,375	258,000
2021–09–22	1.137948	0.0125	−46,875	211,125
2021–09–23	1.130630	−0.0073	27,375	238,500
2021–09–24	1.131484	–	–	–

Table 12 Sensitivity analysis for different spot rate with full hedging at expiry date[2]

Spot price	Cash flow at spot rate ($)	Cash flow with full hedging ($)	Total gain/loss ($)
1.11	4,162,500.0	4,387,500.0	225,000.0
1.12	4,200,000.0	4,387,500.0	187,500.0
1.13	4,237,500.0	4,387,500.0	150,000.0
1.14	4,275,000.0	4,387,500.0	112,500.0
1.15	4,312,500.0	4,387,500.0	75,000.0
1.16	4,350,000.0	4,387,500.0	37,500.0
1.17	4,387,500.0	4,387,500.0	0.0
1.18	4,425,000.0	4,387,500.0	−37,500.0
1.19	4,462,500.0	4,387,500.0	−75,000.0
1.20	4,500,000.0	4,387,500.0	−112,500.0
1.22	4,575,000.0	4,387,500.0	−187,500.0
1.24	4,650,000.0	4,387,500.0	−262,500.0

based on current futures price. Simulated price is 1.170292 on 2021–07-02 which has increased by 0.000292 compared to the day before. The gain/loss on day 2021–07-02 is calculated as: $125000 * (1.170292 - 1.17) * 30 =$ \$1095, where 125,000 is contract size and 30 is number of contracts. The amount \$1095 is subtracted from \$90,000 since the company is in put position.

The scenario analysis is for full hedging and different rate of hedge can be obtained. The company may gain or loss at expiry date, and they therefore may not take full hedge position.

6.2.3 Sensitivity Analysis of Spot Rate

On day of expiry, spot rate may take different values than expected. At this point, companies may create a sensitivity analysis table to see possible gains or losses. In Table 12, total gain/losses are given for different spot prices at expiry.

Since the company takes short position in Euro futures, they make gains for any spot price less than 1.17 and vice versa. The total gain/losses of company is illustrated in Fig. 27.

Another sensitivity analysis is made for hedge ratios, and gain/losses is given for different hedge ratios in Table 13.

Total cash flows under different scenarios for hedge ratios are illustrated in Fig. 28.

In this case, a hedging method is used to stabilize the financial position. A company aims to hedge their cash flow on a future date. A margin account is created, and marking to market process is shown in accordance with simulated futures price. Two different sensitivity analyses are also applied for both spot rate and hedge ratio. All

[2] All values given in Table 12 are in US Dollar.

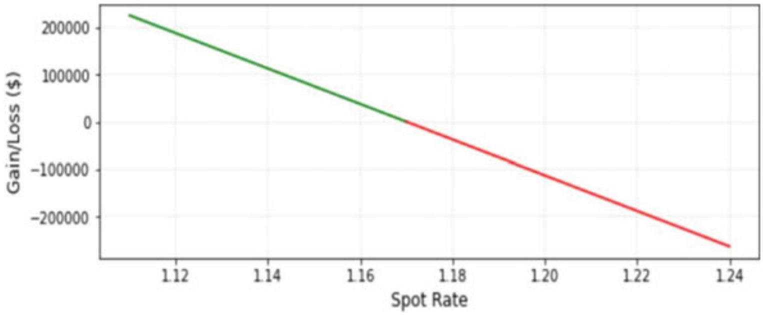

Fig. 27 Gain and losses under different spot rate scenarios

Table 13 Total cash flow under different hedge ratios

Tables hedge ratio (%)	Tables total cash flow ($)
0.0	4,162,500.0
10.0	4,185,000.0
20.0	4,207,500.0
30.0	4,230,000.0
40.0	4,252,500.0
50.0	4,275,000.0
60.0	4,297,500.0
70.0	4,320,000.0
80.0	4,342,500.0
90.0	4,365,000.0

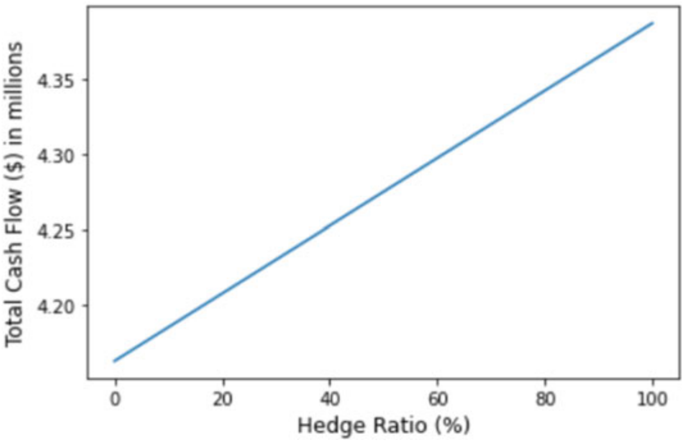

Fig. 28 Total cash flow under different hedge ratios

scenarios on expiry date were summarized in Tables 11, 12 and 13 and Figs. 26, 27 and 28. The optimal hedge ratio can also be determined theoretically but it is preferred to use sensitivity analysis to decide the amount of equity to hedge.

References

1. Tuovila A (2021) Financial analysis. Investopedia. https://www.investopedia.com/terms/f/financial-analysis.asp/. (Access: 23 July 2021)
2. Friedlob GT, Schleifer LL (2003) Essentials of financial analysis. Wiley p 23
3. Kumawat D (2020) An introduction to financial analysis, analytic steps. https://www.analyticssteps.com/blogs/introduction-financial-analysis/. (Access: 23 July 2021)
4. Fridson MS, Alvarez F (2011) Financial statement analysis: a practitioner's guide. Wiley p 597
5. Approved prep provider CFA institute: financial reporting and analysis (2021). https://ift.world/booklets/fra-understanding-cash-flow-statements-part1/. (Access: 23 July 2021)
6. Elliott JW, Uphoff HL (1972) Predicting the near term profit and loss statement with an econometric model: a feasibility study. J Account Res 259–274
7. Barnes P (1987) The analysis and use of financial ratios. J Bus Finance dan Acc 14(4):449
8. Chen KH, Shimerda TA (1981) An empirical analysis of useful financial ratios. Financ Manage 51–60
9. Rao NV, Atmanathan G, Shankar M, Ramesh S (2013) Analysis of bankruptcy prediction models and their effectiveness: an Indian perspective. Gt Lakes Her 7(2)
10. Url: https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data)
11. Musil CM, Warner CB, Yobas PK, Jones SL (2002) A comparison of imputation techniques for handling missing data. West J Nurs Res 24(7):815–829
12. Sarmas E, Xidonas P, Doukas H (2020) Multicriteria portfolio construction with python. Springer International Publishing, pp 1–3. https://doi.org/10.1007/978-3-030-53743-2_1
13. Markowitz H (1952) Portfolio selection. J Financ 7:77–91
14. Sharpe WF (1964) Capital asset prices: a theory of market equilibrium under conditions of risk. J Financ 19:425–442. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x
15. Lintner J (1965) Security prices, risk, and maximal gains from diversification. J Financ 20:587–615. https://doi.org/10.1111/j.1540-6261.1965.tb02930.x
16. Mossin J (1966) Equilibrium in a capital asset market. Econometrica 34(4):768–783. https://doi.org/10.2307/1910098
17. Fama EF, French KR (1983) Common risk factors in the returns on stocks and bonds. J Financ Econ 33(1):3–56. https://doi.org/10.1016/0304-405X(93)90023-5
18. Romero PJ, Balch T (2014) Modern portfolio theory: the efficient frontier and portfolio optimization. What hedge funds really do: an introduction to portfolio management. Business Expert Press. pp 71–81
19. Black F, Litterman R (1991) Combining investor views with market equilibrium. J Fixed Income 1(2):7–18. https://doi.org/10.3905/jfi.1991.408013
20. Hull J (2018) Options, futures, and other derivatives, 10th edn. Pearson
21. CME Group (2021) Retrieved 1 September 2021, from https://www.cmegroup.com/markets/products.html#sortAsc&sortField
22. Endel F, Piringer H (2015) Data wrangling: making data useful again. IFAC-PapersOnLine 48(1):111–112
23. Furche T, Gottlob G, Libkin L, Orsi G, Paton NW (2016) Data wrangling for big data: challenges and opportunities. EDBT 16:473–478
24. Field A (2009) Logistic regression. Discovering Stat Using SPSS 264:315
25. Yacouby R, Axman D (2020) Probabilistic extension of precision, recall, and F1 score for more thorough evaluation of classification models. In: Proceedings of the first workshop on evaluation and comparison of NLP systems. pp 79–91

Mahmut Sami Sivri is currently a lecturer at Industrial Engineering Department at Istanbul Technical University. He is also working on some R&D projects as Director of Data Analytics at ITU's Technopark. He received the B.S. degree in Computer Engineering and the M.Sc. degree in Engineering Management from Istanbul Technical University. He worked in various companies and positions in the IT industry since 2008. His current research interests include machine learning, sentiment analysis, deep learning, big data and its applications, Industry 4.0, financial technologies, data analytics, supply chain and logistics optimization.

Abdullah Emin Kazdaloglu is a research assistant and a Ph.D. student in industrial engineering department at Istanbul Technical University (ITU). He received his B.Sc. degree in management engineering from ITU in January of 2018 and also received his M.Sc. degree in industrial engineering from ITU in August of 2021. He has a passion for data science, machine learning and artificial intelligence, and his current research interests are statistics, data science and business analytics methods. He is eager to develop his professional skills and curious to discover and implement new manners, technologies and tools especially in field of machine learning, statistic and computer science ecosystems, data analytics and business knowledge.

Emre Ari is a Research Assistant and a Ph.D. student in the Department of Industrial Engineering at Istanbul Technical University, Turkey. In 2016, he received his M.Sc. Degree in Statistics from Queen Mary University of London in UK and he received his B.Sc. degree in Mathematics from Kahramanmaraş Sütçü İmam University in 2009. His current research interests lie in the area machine learning, deep learning, reinforcement learning, data analysis, financial approaches and statistics. He plans to continue his academic career by trying to find new approaches in his interest areas.

Hidayet Beyhan is a Research Assistant and a Ph.D. Candidate in the Management Program at Istanbul Technical University, Turkey. He received his M.Sc. degree in Accounting and Finance from Swansea University and B.Sc. degree in Mathematics from Eskişehir Osmangazi University. His current research interests are agent-based modelling in finance, quantitative finance and computational finance. He continues his research in multi-agent financial markets and plans to extend his research with the application of machine learning methods in finance.

Alp Ustundag is the Head of Industrial Engineering Department of Istanbul Technical University (ITU) and the coordinator of M.Sc. in Big Data & Business Analytics Program. He is also the CEO of Navimod Business Analytics Solutions located in ITU Technopark (http://navimod.com/). He has worked in IT and finance industry from 2000 to 2004. He continued his research studies at the University of Dortmund between 2007 and 2008 and completed his doctorate at ITU in 2008. He has conducted a lot of research and consulting projects in the finance, retail, manufacturing, energy and logistics sectors. His current research interests include artificial intelligence, data science, machine learning, financial and supply chain analytics. He has published many papers in international journals and presented various studies at national and international conferences.