



Fundamentals of Business Management - Project

Carella A., Cattari S., Jallow E., Poiani M.

A.Y. 2023/2024

Contents

1	Scope and Approach	1
1.1	Scope	1
1.2	Approach	1
1.2.1	Initial setting	1
1.2.2	Elaboration of the problem	2
1.2.3	Final solution	3
2	Selection of the most important items from the P&L	4
2.1	Profit & Loss statement	4
2.1.1	Relevant items in our project	7
3	Choice of the most relevant macroeconomic values	9
3.1	Where we have found the values	9
3.2	Macroeconomic values used in our project	9
4	Overview of technical solution	11
4.1	Implementation	11
5	Technical details	14
5.1	Introduction	14
5.2	Model Creation	14
5.3	Model Evaluation	14
5.4	Frontend Integration	15
5.5	Libraries	15
6	Mock-Up and Demo	16
6.1	Landing page	16
6.2	Evaluation and Selection of Model	16
6.3	Dashboard	17
6.3.1	Dasboard items	17
6.4	What-if Analysis	19
6.5	Responsive Site	20

Chapter 1

Scope and Approach

1.1 Scope

In the following pages, we elucidate the process undertaken to deliver a solution for the Business Data Challenge presented by CCH Tagetik. Our project, titled "Forecast in a Bank/Finance Sector," addressed the specific request: "A company in the banking sector wants to incorporate the Net Inflow forecast as a starting point for its 2023 budget, based on macro-economic data and operational/promotion costs to perform margin simulations and react to adverse scenarios." Our goal was to present a mock-up of a dashboard featuring key performance indicators (KPIs) essential for conducting what-if analyses post the forecasted results. Furthermore, we aimed to provide users with the capability to verify how alterations in the macroeconomic scenario could impact the EBIT, along with various assumptions about the business. Incorporating macroeconomic values in the what-if analysis is worthwhile, even if these values are beyond the control of financial intermediaries. This inclusion enables the forecasting of EBIT in optimistic, pessimistic, and normal macroeconomic scenarios. To better assess the impact of significant macroeconomic events, the data used for our model spans from 2018, the biennium preceding the COVID-19 pandemic, until 2022. To execute the task effectively, we divided it into three phases. Initially, we curated the most pertinent and valuable items from the Profit and Loss statement (P&L), as well as selected macroeconomic values, to form our data set. Subsequently, we utilised the data set to train the chosen machine learning model. Finally, upon confirming the accuracy of the forecast, we transmitted the values to the dashboard for visualisation and execution of the what-if analysis.

1.2 Approach

1.2.1 Initial setting

In this section, we describe the phases from the beginning to the end of the project. As we can see from the table 1.1 in the team we were four students and each of us had a specific task. Despite the divisions, we cooperate synergetically to provide the best solution.

Table 1.1: Team #2 Components:

Surname	Name	Responsibilities
Poiani	Marco	Choice of the most significant P&L items as collection drivers (Team Leader)
Cattari	Simona	Choice of the most significant macroeconomic variables
Carella	Alessandro	Preparation and choice of one or more ML models for the train and forecast phases
Jallow	Ebrima	Choice of output evaluation metrics

To complete the initial setting and start with the elaboration of the project we chose a name and designed a logo for our team. The name of our team is **CPCJ**, derived from the initial of our surname, and for the logo, we opted for the wordmark style, as we can see from the figure 1.1 below.



Figure 1.1: Team's Logo

1.2.2 Elaboration of the problem

The initial phase of the project involved collecting relevant data to construct our dataset, crucial for training the machine learning model and generating forecasts. Our focus was on the Italian market, leading us to reference the financial statements of the top five banks in Italy by assets: Unicredit, Banco BPM, Monte dei Paschi di Siena (MPS), Intesa San Paolo, and BPER. As mentioned earlier, we utilized financial statements spanning from 2018 to 2022 to assess the impact of adverse conditions, specifically those arising from the COVID-19 pandemic. The resulting dataset comprises 25 rows and 50 columns. We deemed this volume of data sufficient for training our model and addressing the defined problem. However, recognizing the potential to enhance forecast reliability, we acknowledge the possibility of expanding the dataset by including more banks, additional years, or additional features. Having identified key elements from the banks' income statements, we proceeded to incorporate relevant macroeconomic indicators that could significantly influence year-end net income. Our analysis included the following values: GDP Index, Unemployment rate, Producer Price Index (PPI), Exchange Rate €/€, Consumer Price Index (CPI), COVID-19 Stringency Index, and the Real Interest Rate. We delve into these aspects of the income statement and macroeconomic values in the dedicated chapter 2 and chapter 3. Finally,

since the costs of employees are generally high and their weight in the operational costs is relevant, we include two columns in our dataset about the number of employees and the average cost per employee. Once terminated the business-oriented part, we entered the technical phase, consisting of two tasks: selecting and training the machine learning model and developing a dashboard. The dashboard was crafted using React and integrates the model. We prioritized user-friendliness in the design, ensuring a minimalistic yet effective interface that enables users to visualize data, observe trends, and conduct what-if analyses. All these concepts will be well-explained in each of the following chapters. Following the initial allocation of responsibilities, we established a weekly meeting to collaboratively review individual accomplishments from the preceding week and strategically plan micro-tasks for the upcoming week. This systematic approach ensured equal participation from all team members, fostering comprehensive awareness of the project's progress among the entire team.

1.2.3 Final solution

The final solution involved purely the dashboard made up of three pages, from which the user can conduct all the analyses requested from the assignment. We highlight that this dashboard is generic and not tailor-made for a specific client and its needs, but we may easily modify it to come across the specific requests of the user. The business user has the possibility to interact only with the offered tools, but we might provide the documentation to allow the auto-maintenance and update of the dashboard, destined to technical departments. We reported the dashboard and its explanation in the last chapter of this document.

Chapter 2

Selection of the most important items from the P&L

2.1 Profit & Loss statement

The first task of our project was to extract valuable data from the financial statements of the selected banks to build our dataset. The problem required forecasting the Net inflow of the following year. Being the term Net Inflow related to the Cash Flow statement, initially, we hypnotized finding a way to predict this value. Considering the requests and talking with mentors we understood that the real objective was not predicting the Net cash Inflow but the Net Income, a value reported at the bottom of the Income statement, also known as Profit & Loss statement. The Net income is a value calculated through this financial statement, that measures the company's profitability over a specific period, but it is also reported in the balance sheet, incorporated in the item "Retained earnings" since the three financial statements are linked among them. For our objective, we need to figure out the reasons behind the Net income, by analyzing multiple items in the income statement. Consequently, we confirmed our choice of using the Income statements because it was the most appropriate document for our task. During the year financial institutions publish the financial statements quarterly, but we used the comprehensive one published at the end of the year. Additionally, banks generically are groups and for this reason, they must prepare a consolidated Income statement. The key distinction between the income statement and the consolidated version is that the first one includes only the financial results of a single company, while the second one incorporates the results of both the parent company and its subsidiaries. Consolidated financial statements are especially important for groups of companies where the parent has controlling ownership interests in other entities, the subsidiaries. It provides a more accurate representation of the overall financial health and performance of the entire group. We took the income statement directly from the annual report published on the websites of the financial institutions [2, 3, 4, 5, 6]. Initially, we tried to gather data from the online database Aida, but we noticed a consistent lack of relevant information. Using the reports provided by Aida we had the final results about Net Income or some KPIs, but we didn't have all the data that composed them. The interesting part of the complete financial reports was that, besides financial reports, they provided a lot of useful information and insight about

CHAPTER 2. SELECTION OF THE MOST IMPORTANT ITEMS FROM THE P&L

the company and even data about the global and local macroeconomic scenario at the end of the year. We considered all of these values precious for our prediction, since the larger the amount of valuable data, the more reliable the forecast. Below we reported, for example, one of the 2022 income statements from which we selected data. The one under consideration belongs to Unicredit.

ITEMS	YEAR	
	2022	2021
10. Interest income and similar revenues	16,339	12,703
of which: interest income calculated with the effective interest method	13,426	9,846
20. Interest expenses and similar charges	(5,715)	(3,612)
30. Net interest margin	10,624	9,091
40. Fees and commissions income	8,105	7,963
50. Fees and commissions expenses	(1,418)	(1,260)
60. Net fees and commissions	6,687	6,703
70. Dividend income and similar revenues	437	351
80. Net gains (losses) on trading	859	1,472
90. Net gains (losses) on hedge accounting	367	49
100. Gains (Losses) on disposal and repurchase of:	457	244
a) financial assets at amortised cost	133	53
b) financial assets at fair value through other comprehensive income	133	141
c) financial liabilities	191	50
110. Net gains (losses) on other financial assets/liabilities at fair value through profit or loss:	563	(469)
a) financial assets/liabilities designated at fair value	1,139	(306)
b) other financial assets mandatorily at fair value	(576)	(163)
120. Operating income	19,994	17,441
130. Net losses/recoveries on credit impairment relating to:	(2,061)	(1,648)
a) financial assets at amortised cost	(2,031)	(1,630)
b) financial assets at fair value through other comprehensive income	(30)	(18)
140. Gains/Losses from contractual changes with no cancellations	(3)	(5)
150. Net profit from financial activities	17,930	15,788
160. Net premiums	-	-
170. Other net insurance income/expenses	-	-
180. Net profit from financial and insurance activities	17,930	15,788
190. Administrative expenses:	(10,302)	(11,257)
a) staff costs	(6,208)	(7,045)
b) other administrative expenses	(4,094)	(4,212)
200. Net provisions for risks and charges:	33	(377)
a) commitments and financial guarantees given	42	(26)
b) other net provisions	(9)	(351)
210. Net value adjustments/write-backs on property, plant and equipment	(764)	(850)
220. Net value adjustments/write-backs on intangible assets	(550)	(621)
230. Other operating expenses/income	601	566
240. Operating costs	(10,982)	(12,539)
250. Gains (Losses) of equity investments	297	(1,462)
260. Net gains (losses) on property, plant and equipment and intangible assets measured at fair value	11	(19)
270. Goodwill impairment	-	-
280. Gains (Losses) on disposals on investments	33	11
290. Profit (Loss) before tax from continuing operations	7,289	1,779
300. Tax expenses (income) of the year from continuing operations	(819)	343
310. Profit (Loss) after tax from continuing operations	6,470	2,122
320. Profit (Loss) after tax from discontinued operations	3	4
330. Profit (Loss) of the year	6,473	2,126
340. Minority profit (loss) of the year	(15)	(30)
350. Parent Company's profit (loss) of the year	6,458	2,096

Figure 2.1: Unicredit Income Statement 2022

In our dataset, we reported the values of all the rows up to the item "330. Profit (Loss) of the year". Generically, banks' income statements are equal because they compile the financial statement following the dispositions of the Bank of Italy, also if there are a few slight differences from one institution to another about less relevant items. Considering the income statement, from the first item "10. Interest income and similar revenues" up to the item "140. Gains/Losses from contractual changes with no cancellations" we found the

elements that make up the earnings of the bank. Then, item "150. Net profit from financial activities" provides the sum of all these elements. The items from "190. Administrative expenses" till "230. Other operating expenses/income" comprised the costs borne by the bank, we encounter the sum of all these costs in item "240. Operating Cost". Finally, we found the EBIT (Earnings Before Interest and taxes) and the Net income, respectively as "290. Profit (Loss) before tax from continuing operations" and "330. Profit (Loss) of the year". Being the Income statement reported a consolidated version, we can notice the item "350. Parent company's profit(Loss) of the year" too. All the items presented so far are in turn made up of multiple values and variables. For instance, if we break down item 10 of the income statement we discover that it is the result of several elements, as illustrated in the figure 2.2 below.

ITEMS/TYPES	YEAR 2022				YEAR 2021
	DEBT SECURITIES	LOANS	OTHER TRANSACTIONS	TOTAL	TOTAL
1. Financial assets at fair value through profit or loss	246	102	845	1,193	943
1.1 Financial assets held for trading	158	12	845	1,015	754
1.2 Financial assets designated at fair value	2	-	-	2	3
1.3 Other financial assets mandatorily at fair value	86	90	-	176	186
2. Financial assets at fair value through other comprehensive income	747	-	X	747	723
3. Financial assets at amortised cost	780	11,899	X	12,679	9,123
3.1 Loans and advances to banks	101	1,682	X	1,783	373
3.2 Loans and advances to customers	679	10,217	X	10,896	8,750
4. Hedging derivatives	X	X	278	278	(72)
5. Other assets	X	X	468	468	362
6. Financial liabilities	X	X	X	974	1,624
Total	1,773	12,001	1,591	16,339	12,703
of which: interest income on impaired financial assets	2	331	-	333	383
of which: interest income on financial lease	X	463	X	463	455

Figure 2.2: Unicredit 2022 Income Statement: 10. Interest income and similar revenues breakdown

We emphasize this aspect because it is crucial for the proper usage of what-if analysis. This is important because users have the flexibility to perform analyses by adjusting either a broader item or a more specific one. In the first scenario, users can observe the effects of increasing or decreasing a particular item within the Income Statement (e.g."10. Interest income and similar revenues" or "30. Net Interest margin" Fig:2.1). In the second scenario, users can assess the impact of changing the value of an element embedded in one of the items of the Income Statement (e.g."1.1. Financial assets held for trading" Fig:2.2). Since our dashboard is designed to be more generic, we provide users with the option to conduct what-if analyses using values of the Income Statement. Below we present and analyze the income statement items that we considered more relevant for our project, chosen because are broader and reflect better the financial health of the financial institutions. Although we consider some values more significant for our task, we decided to use all of the items of the income statement for training the model, in order to obtain a more reliable forecast.

2.1.1 Relevant items in our project

- **Net interest margin:** This particular item serves as the primary income source for all the financial institutions under scrutiny. Recognizing its significance, we have incorporated it into our analysis. Moreover, we offer users the capability to observe how alterations in this value can influence the EBIT. Once the user has assessed the impact of this item through the what-if analysis, he can take direct action on one or more of the sources contributing to the net interest margin, enabling them to effectively generate improvements.
- **Net fees and commissions:** This item, along with the Net interest margin, composes almost the total operating income. It is made up of two main elements "fees and commissions income" and "fees and commissions expenses". Even in this case, a tiny change in the value can have a considerable impact on the EBIT, consequently, we offered the possibility of handling it during the what-if analysis.
- **Operating income:** We consider this item valuable because, besides the information provided for the forecast of the Net income, it gives an important insight about profits. Additionally, it is not the mere sum of the previous two items, but it includes the dividend income, Net gains(Losses) on trading and other minor values. This value, as the previous two, can be modified during the what-if analysis.
- **Net profit from financial activities:** We consider this item insightful because provides a comprehensive value. However, it is less relevant than others for the what-if analysis because the user cannot clearly understand which are the elements impacted by the analysis.
- **Administrative expenses:** This item is the first related to the expenses. Banks work in the services industry, consequently, the administrative expenses are considerably high. For this reason, we incorporate this value in our dataset such as in our dashboard to conduct the what-if analysis using it. Almost the totality of the expenses falls in this item, which includes "staff costs" and "other administrative expenses".
- **Operating costs:** This item provides a comprehensive view of the costs incurred by the bank. The promotion costs mentioned in the delivery of the project are allocated here, but observing various financial reports they are not relevant to the final result of the variable under consideration. For instance, the promotion costs for the income statement reported above are 1% of the operating costs, while the expenses for the staff are 57% of the total. An example of a more determinant expense is represented by the cost of "Information and communication technology", which compose 11% of the operating costs, as we can see in figure 2.3. Similar to our previous discussion on the Net profit from financial activities, in this case as well, we acknowledge that this item may not yield significant insights for the what-if analysis. However, it contributes a substantial value to our model.

TYPE OF EXPENSES/SECTORS	€ million	
	YEAR 2022	YEAR 2021
1) Indirect taxes and duties	(596)	(568)
1a. Settled	(596)	(566)
1b. Unsettled	-	(2)
2) Contributions to Resolution Funds and Deposit Guarantee Schemes (DGS)	(901)	(910)
3) Guarantee fee for DTA conversion	(104)	(104)
4) Miscellaneous costs and expenses	(2,493)	(2,630)
a) Advertising marketing and communication	(116)	(153)
b) Expenses relating to credit risk	(84)	(119)
c) Indirect expenses relating to personnel	(68)	(55)
d) Information & Communication Technology expenses	(1,175)	(1,142)
Lease of ICT equipment and software	(74)	(73)
Software expenses: lease and maintenance	(322)	(270)
ICT communication systems	(67)	(72)
Services ICT in outsourcing	(588)	(605)
Financial information providers	(124)	(122)
e) Consulting and professionals services	(114)	(182)
Consulting	(75)	(136)
Legal expenses	(39)	(46)
f) Real estate expenses	(394)	(405)
Premises rentals	(47)	(48)
Utilities	(140)	(132)
Other real estate expenses	(207)	(225)
g) Operating costs	(542)	(574)
Surveillance and security services	(48)	(59)
Money counting services and transport	(51)	(49)
Printing and stationery	(33)	(27)
Postage and transport of documents	(66)	(75)
Administrative and logistic services	(174)	(186)
Insurance	(64)	(68)
Association dues and fees and contributions to the administrative expenses deposit guarantee funds	(69)	(65)
Other administrative expenses - other	(37)	(45)
Total (1+2+3+4)	(4,094)	(4,212)

Figure 2.3: Unicredit 2022 Income Statement: 190. Other administrative expenses: breakdown

- **EBIT:** This value is the pillar of the project because we predict it and then, through the what-if analysis, we always observe the impact on it. It indicates the financial result of the business before interests and taxes.
- **Net income:** This element is one of the most important because the project starts from here since we have to incorporate the forecast of the net income as a starting point for the following year, in our case for 2023. It represents the financial result after interest and taxes.

In addition to this list of KPIs traceable in the Income statement, there are many other useful KPIs for a bank, such as the Net Profit margin, the ROA, and the ROE. The first one is a percentage that represents the profitability of a company in relation to its revenue. It is a measure of how much profit a company makes for every dollar of revenue it generates. The second indicator is a measure of a bank's overall profitability. It is calculated by dividing net income by average total assets, while the third value is a measure of a bank's profitability about its shareholders' investment. It is calculated by dividing net income by average shareholders' equity. Banks can manage the values mentioned above to varying extents. However, other values beyond their direct control can significantly influence the net income at the end of the year. These are the macroeconomic factors, which we will delve into extensively in the next chapter.

Chapter 3

Choice of the most relevant macroeconomic values

3.1 Where we have found the values

We found the values of macroeconomic variables in the official financial reports of the banks under examination and on the site of the World Bank, under the section "Data".

The values of macroeconomic variables are present in the official financial reports of banks for various reasons. Banks are required to maintain a certain level of transparency for their shareholders and the general public to openly communicate the financial and operational conditions of the institution. They have regulatory obligations that require the disclosure of certain financial and macroeconomic data; these regulations are designed for financial stability and transparency. Investors (shareholders and potential securities buyers) seek comprehensive information to make informed decisions. Banks, therefore, provide relevant macroeconomic data as part of their efforts to communicate effectively with investors. Financial analysts and rating agencies regularly examine the macroeconomic data of banks to assess their financial strength and ability to manage risks. The publication of such information facilitates the assessment process. Providing accurate and timely information on macroeconomic indicators contributes to the credibility and reputation of a bank. Transparency in this context can positively influence the perception of the public and stakeholders. Best practices in the banking sector often require the disclosure of key information. Banks that follow these practices can be considered more reliable and professional.

On the other hand, the World Bank is an international organization founded in 1945 with the initial goal of reconstructing Europe and Japan after World War II. Over time, it has expanded its role, focusing on the economic development of developing countries. It collects and publishes data reflecting global economic and social conditions:

3.2 Macroeconomic values used in our project

In this section, we will explore key macroeconomic indicators that play a significant role in shaping the economic landscape. Understanding these indicators is crucial for analyzing the potential impact on the banking sector. The following indices provide insights into the

overall health of an economy and its various components.

- **Gross Domestic Product (GDP):** Plays a crucial role in reflecting the current overall health of a nation's economy. Fluctuations in the GDP, representing the total value of goods and services produced, can significantly impact the demand for financial services and the risk-taking attitude of economic players. A consistently growing GDP can be seen as an indicator of a favourable environment for investments and access to credit, whereas a contraction may suggest the opposite.
- **Unemployment Rate:** Indicates the percentage of individuals without jobs compared to the total workforce, holding a central role in analyzing the stability of the labour market. Its variations can influence the demand for loans, shape consumer confidence, and impact the real estate sector. A high unemployment rate could signal a reduced ability to repay loans and a consequent decline in the demand for financial services.
- **Producer Price Index (PPI):** An indicator reflecting the changing dynamics of production costs plays a crucial role in influencing profit margins in the banking sector. An increase in production costs can put pressure on profit margins, while a decrease could contribute to improving the profitability of the sector.
- **Consumer Price Index (CPI):** Represents inflation and monitors changes in prices of goods and services consumed. It plays a crucial role in influencing real interest rates and the overall demand for financial services. A high inflation rate can erode the purchasing power of the currency.
- **Exchanges rates € / \$:** Influence the relative value of currencies. They have a deep impact on international trade, import-export operations, and the activities of multinational companies. Fluctuations in exchange rates can affect profits from foreign currency activities and the ability of a company to compete effectively on a global scale.
- **Real interest rate:** Considering the effect of inflation on nominal interest rates, plays a decisive role in shaping incentives for investments and borrowing. A negative real interest rate can encourage investments and access to credit, while a positive rate may have the opposite effect.
- **Covid-19 Stringency Index:** Reflects the degree of limitations and restrictions related to the pandemic. This index can influence economic mobility and business activities, with potential direct effects on sectors such as loans, investments, and business operations.

Chapter 4

Overview of technical solution

4.1 Implementation

Our team has endeavoured to create a forecasting system aimed at predicting net inflow and conducting what-if analyses in the financial domain. To achieve this, we utilized Python along with the scikit-learn library, leveraging their capabilities in scientific computing and machine learning. In our development process, we opted for Python due to its flexibility and the availability of extensive libraries. The scikit-learn library, a valuable resource for machine learning in Python, played a key role in building, training, and validating our predictive models. This approach aims to enhance the accuracy and reliability of our predictions within the dynamic landscape of financial scenarios. For the user interface, we opted for Vite on the frontend, coupled with Material UI for a modern and intuitive design. This combination provides a seamless and responsive experience for users interacting with our forecasting system. Vite, known for its speed and efficiency, enhances the overall performance of our frontend application. On the backend, we implemented Django to host the model files and establish a reliable infrastructure. Django's versatility and scalability make it an ideal choice for handling the complexities of hosting Machine Learning models. We created a robust API endpoint that allows seamless communication between the frontend and backend. This API endpoint efficiently processes requests from the frontend triggers the model predictions and returns the results promptly. The models we developed not only forecast net inflow but also empower users to perform what-if analyses. This feature enables users and stakeholders to explore various hypothetical scenarios and understand the potential outcomes. The integration of this capability adds a layer of strategic decision-making to our forecasting system. In summary, our end-to-end solution seamlessly combines the strengths of Python and scikit-learn for meticulous model development, integrates Vite and Material UI to offer an intuitive frontend experience, and relies on Django for a steadfast backend infrastructure. This synergistic approach culminates in a forecasting system that not only delivers predictions but also empowers users with dynamic what-if analyses for informed decision-making.

The data flow in our system follows a structured path from the user's end to the backend, with various types of data being processed and returned to the frontend seamlessly. The illustration below outlines the key steps in this data flow 4.1:

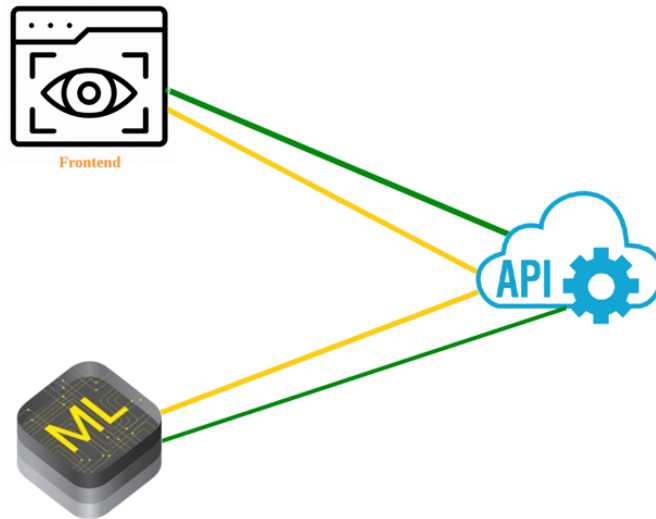


Figure 4.1: How data flow in our system

- **User Dataset Upload:** Users initiate the process by uploading a dataset to the frontend. This dataset serves as the foundation for subsequent predictions and what-if analyses.
- **Backend Processing:** The uploaded dataset is sent to the backend, where it undergoes preprocessing. During this phase, mean values for specific features which are crucial for the prediction are calculated. These mean values are needed for initializing sliders and inputs for the what-if analysis.
- **Initialization Data to Frontend:** The calculated mean values are then transmitted back to the frontend. This data serves as the default or initialized values for sliders and inputs in the user interface, providing users with a starting point for their what-if analyses.
- **User Interaction and Input:** Users interact with the frontend, adjusting sliders and input parameters based on their specific scenarios of interest. These inputs are then sent back to the backend and forwarded to the model for further processing.
- **Prediction and What-If Analysis:** Leveraging the initialized values and user inputs, the backend performs predictions and what-if analyses using the machine learning models developed with Python and scikit-learn.
- **Results to Frontend:** The results of the predictions and what-if analyses are sent back to the frontend via API. Users can visualize and analyze the outcomes of their chosen scenarios.
- **Interactive User Experience:** The frontend, powered by Vite and Material UI, provides an interactive and responsive user experience. Users can seamlessly navigate through the application, adjust parameters, and view real-time updates of predictions and analyses.

- **API Endpoint for Communication:** Django facilitates communication between the frontend and backend through a well-defined API endpoint. This ensures a smooth exchange of data, enabling a dynamic and efficient user experience.

By orchestrating this comprehensive data flow, our system enables users to upload datasets, receive initialized values, interact with a user-friendly interface, and obtain real-time predictions and what-if analyses, all within a seamless and efficient infrastructure. This approach enhances user engagement and facilitates informed decision-making in diverse scenarios.

Chapter 5

Technical details

5.1 Introduction

This document provides an in-depth exploration of the technical intricacies involved in the Fundamentals of Business Management project. The project involves the development, evaluation, and deployment of prediction models.

5.2 Model Creation

The main file in the project is dedicated to the creation of prediction models. Central to this task is the utilization of scikit-learn, a powerful machine learning library in Python. The project employs two distinct types of models: `DecisionTreeRegressor` and an array of linear regression models including `LinearRegression`, `Ridge`, `Lasso`, and `ElasticNet`. The selection of these models is related to their known efficacy in handling small datasets.

The heart of the modeling process is a 49-feature (same as the previously discussed variables extracted from the banks statements), 25-samples dataset. The models, once trained, are preserved for future use through the serialization mechanism provided by Python's `pickle` module. Additionally, two other essential objects are saved as pickles for future use: the scaler object (constructed with `StandardScaler` from `sklearn.preprocessing`) and the dataset split for evaluation (implemented with `train_test_split` from `sklearn.model_selection`). These objects contribute to the overall reproducibility and scalability of the project. The scaler object is also useful to make predictions on the new data provided by the user in the interface.

5.3 Model Evaluation

The second main file in the project is tasked with the rigorous evaluation of the constructed prediction models. A range of performance metrics is computed for each model, providing a comprehensive assessment of their effectiveness. These metrics include mean squared error, mean absolute error, R^2 score, explained variance score, accuracy, precision, recall, and F1 score. The results are stored in a file, facilitating easy retrieval for display on the project's

graphical user interface. All the methods involved in the computation of those metrics are implemented by the `sklearn.metrics` library.

5.4 Frontend Integration

The final main file is dedicated to the seamless integration of the project with the frontend. It encapsulates methods designed to handle requests originating from the frontend interface. Of particular note is the method responsible for generating new predictions. This method takes as input new values for prediction, the target variables, and optionally a new dataset to perform the prediction on. The method returns the results for the specified query across all the diverse prediction models created during the project's lifecycle. The user is then tasked with choosing the best prediction model from the panel where the metrics of the various prediction models are showed and has the option to choose which predictions are the most suitable for them.

Within the same file, additional methods exist to fetch evaluation metrics for the models and retrieve default values for variables that users can modify within the interface. For instance, the mean of the values in the dataset serves as a default for changeable parameters, ensuring a user-friendly and intuitive experience.

In the same file is also present the function to send the evaluation data to the frontend.

Here it is imperative to highlight that the overall codebase exhibits a robust structure and adherence to best coding practices. However, certain minor elements may require further attention and refinement, based on the level of grasp of the code one wants to have. To facilitate transparency and collaboration, the entire source code is available for inspection on the project's GitHub repository. This repository serves as a comprehensive resource for developers and stakeholders to delve into the intricacies of the codebase, fostering an environment of continuous improvement and collaborative development.

5.5 Libraries

For the development of the project, we used the following libraries:

- **Django** in the version 4.2.7
- **djangorestframework** in the version 3.14.0
- **matplotlib** in the version 3.7.1
- **numpy** in the version 1.24.3
- **pandas** in the version 2.0.2
- **scikit_learn** in the version 1.2.2
- **seaborn** in the version 0.13.0

Chapter 6

Mock-Up and Demo

6.1 Landing page

The landing page, observable in 6.1, of our developed user interface serves as a gateway to a comprehensive and dynamic business analytics system. Notably, users are provided with the option to upload datasets, enabling them to perform predictions that are vital for informed decision-making. At the top right of the interface are two links leading to the "Dashboard" and "Evaluation" pages, each designed to offer distinct functionalities.



Figure 6.1: Landing page

6.2 Evaluation and Selection of Model

The evaluation page, observable in 6.2, of the developed user interface serves as a pivotal component in our project, providing users with a comprehensive analysis of metrics derived from models computed on the previously uploaded dataset. This interface seamlessly integrates data visualization, allowing users to gain insights into the performance of predictive models.



The screenshot shows the 'Evaluation' tab of the CPCJ dashboard. It features a 'Model Evaluation' table with columns for Model, Mean Squared Error, Mean Absolute Error, R-squared, Explained Variance Score, Accuracy, Precision, Recall, F1 Score, and an 'Activate Model' button. The 'Elastic Net' model is highlighted in yellow.

Model	Mean Squared Error	Mean Absolute Error	R-squared	Explained Variance Score	Accuracy	Precision	Recall	F1 Score	Activate Model
Decision Tree	5703296.6	1591.43	0.31	0.48	0.8	1	0.8	0.88	ACTIVATE
RidgeModel	582488.16	617.12	0.85	0.85	0.4	0.4	0.4	0.4	ACTIVATE
Elastic Net	558678.7	662.55	0.89	0.9	0.4	0.4	0.4	0.4	ACTIVATE
Lasso	302263.27	352.24	0.92	0.92	0.8	0.8	0.8	0.8	ACTIVATE
LinearRegression	1078.28	0.75	1078.28	0.75	0.3	0.3	0.3	0.3	ACTIVATE

Figure 6.2: Evaluation Page

6.3 Dashboard

The dashboard helps to visualize the KPIs, the trend of Earnings Before Income Tax (EBIT), Net Income, macroeconomic variables and the predicted values based on the selected model from the evaluation page.

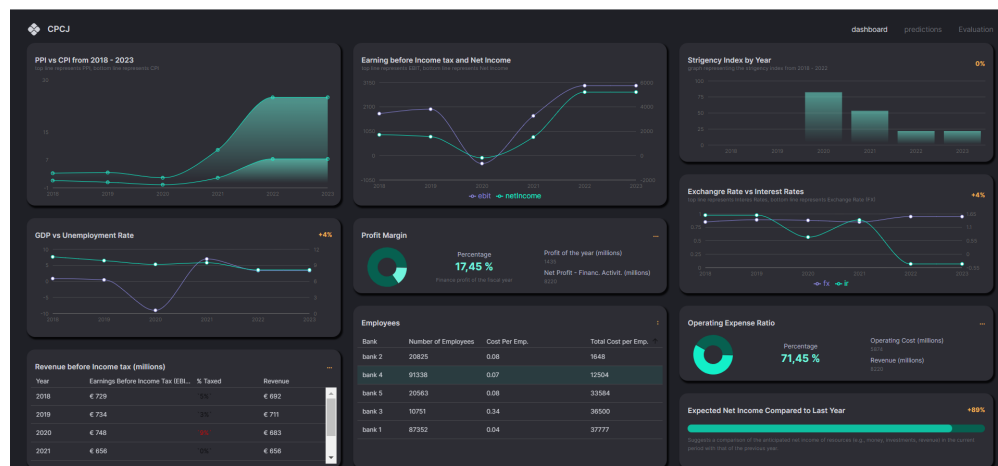


Figure 6.3: Dashboard page

6.3.1 Dashboard items

- **Producer Price Index (PPI) and Consumer Price Index (CPI) Trends:** This item displays a line chart illustrating the trends of Producer Price Index (PPI) and Consumer Price Index (CPI) from 2018 to 2023. The top line represents Net Income, while the bottom line shows CPI.
- **Earnings Before Income Tax (EBIT) and Net Income:** This item features a line chart visualizing the trends in Earnings Before Income Tax (EBIT) and Net Income over the period from 2018 to 2023. The blue line represents the trend of EBIT over

the specified period, reflecting operational profitability before income tax expenses. The green line depicts the trend of Net Income, showcasing overall profitability after accounting for all expenses, including taxes.

- **Summary of Covid Stringency Index by Year:** This item presents a bar chart providing a clear depiction of the Covid Stringency Index from 2018 to 2023, offering insights into global pandemic-related restrictions. Each bar represents the index for a specific year, measuring the strictness of pandemic-related policies and restrictions.
- **GDP and Unemployment Trends:** This item features a line chart comparing the trends of Gross Domestic Product (GDP) and Unemployment rates from 2018 to 2023. The blue line represents the trend of GDP, a key indicator of economic health and productivity, while the green line illustrates the trend of Unemployment rates, providing insights into labour market dynamics.
- **Profit Margin Percentage (Pie Chart):** This item features a pie chart illustrating the Profit Margin percentage, derived by dividing the profit of the year by net profit from financial activities. This visualization serves as a valuable tool for decision-makers to understand the composition of profit margins across different financial activities.
- **Exchange Rates and Interest Rates Trends:** This item showcases a line chart depicting the trends of Exchange Rates and Interest Rates from 2018 to 2023. Interest Rates are presented on top in green, allowing for a simultaneous comparison. The chart aids in understanding the influence of interest rate changes on currency markets and assists in decision-making.
- **EBIT (Millions):** The Earnings Before Income Tax (EBIT) column represents the operational profitability of the company before accounting for income tax expenses. It reflects the core earnings generated from business operations. The percentage of tax is crucial in understanding the tax burden on the company's earnings. This column displays the percentage of income tax applied to the Revenue Before Income Tax. Further, the "Revenue" column presents the total income generated by the company before deductions for expenses and taxes. It provides insights into the overall financial health and performance of the business.
- **Employees Table:** This table provides valuable insights into the workforce and associated costs for various banks in the dataset.
- **Operating Expense Ratio Pie Chart:** The Operating Expense Ratio (OER) Pie Chart visually represents the allocation of operating expenses as a percentage of total revenue.

- **Expected Net Income Compared to Last Year Linear Gauge:** The Expected Net Income compared to Last Year Linear Gauge is a visual representation designed to convey the anticipated net income for the forecast year in comparison to the net income from the previous year

6.4 What-if Analysis

The "What If Analysis" page, observable in 6.4, within our developed user interface, serves as a dynamic and interactive tool for users to explore and manipulate key variables influencing business predictions. This feature empowers users to visualize and modify critical factors, including the GDP index, CPI index, PPI, COVID stringency index, unemployment rate, real interest rate, exchange rates percentage, and cost per employee. Through an intuitive interface, users can seamlessly tweak these variables, observing real-time changes in the projected outcomes. The modified variables are then transmitted to the backend, where our models utilize the uploaded dataset and the new data to generate accurate predictions. This iterative process enables users to simulate various scenarios, fostering a deeper understanding of how alterations in these fundamental parameters impact overall business forecasts. The "What If Analysis" page thus stands as a cornerstone of our project, offering users a practical and insightful platform for strategic decision-making and scenario planning.

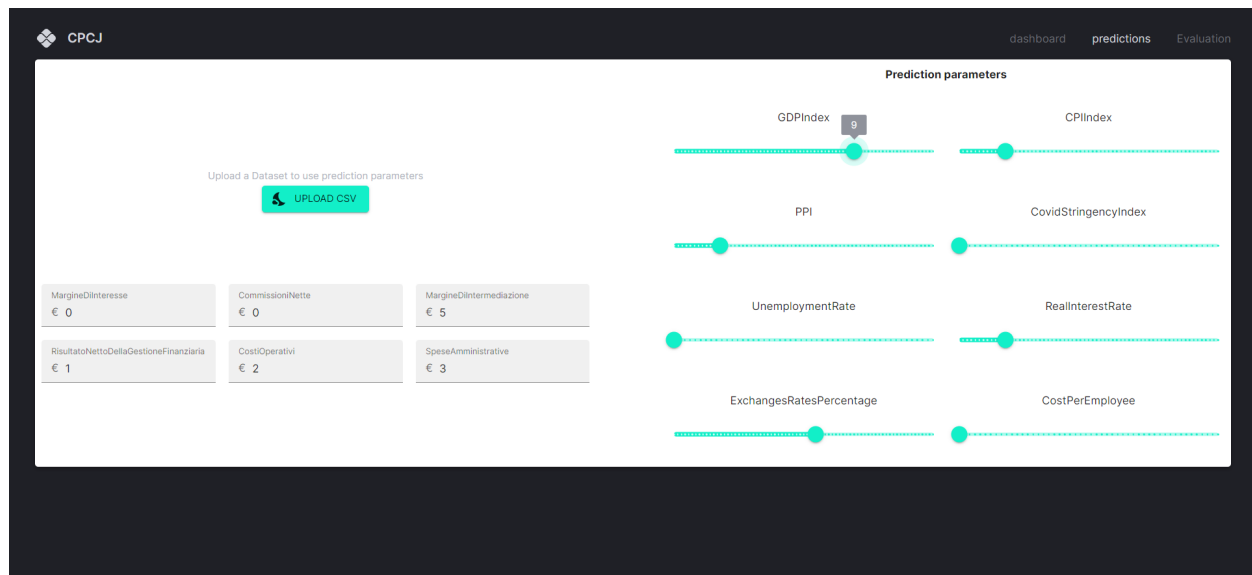


Figure 6.4: What if analysis page (previous Landing page)

6.5 Responsive Site

As we can notice from the picture 6.5 our dashboard is responsive. This means that it works on different devices, from PCs to smartphones and tablets. This functionality allows the user to consult the values in real time. We suggest using the smartphone version only to visualize the dashboard section.

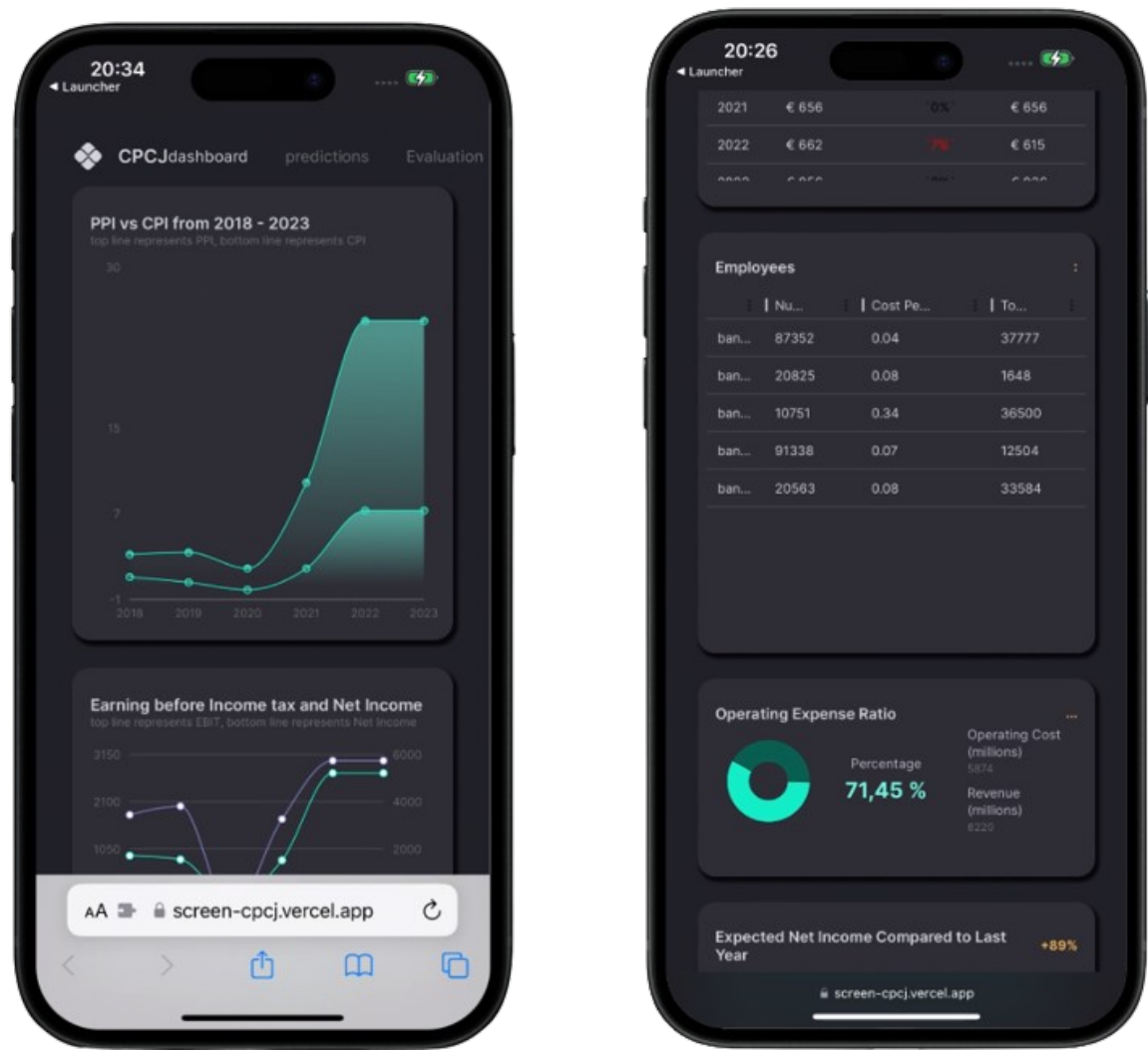


Figure 6.5: Responsive homepage of the dashboard

Bibliography

- [1] The World Bank. *The World Bank*. URL: <https://data.worldbank.org/> (visited on 11/2023).
- [2] BPER. *BPER Financial Reports*. URL: <https://istituzionale.bper.it/investor-relations/bilanci-e-relazioni> (visited on 11/2023).
- [3] Banco BPM. *Banco BPM Financial Reports*. URL: <https://gruppo.bancobpm.it/en/investor-relations/balance-sheets-and-reports/> (visited on 11/2023).
- [4] Monte dei Paschi di Siena. *Monte dei Paschi di Siena Financial Reports*. URL: <https://www.gruppomps.it/en/investor-relations/financial-results/financial-results-year-2022.html> (visited on 11/2023).
- [5] Intesa Sanpaolo. *Intesa Sanpaolo Financial Reports*. URL: <https://group.intesasanpaolo.com/en/investor-relations/financial-reports> (visited on 11/2023).
- [6] Unicredit. *Unicredit Financial Reports*. URL: <https://www.unicreditgroup.eu/en/investors/financial-reporting/financial-reports.html> (visited on 11/2023).