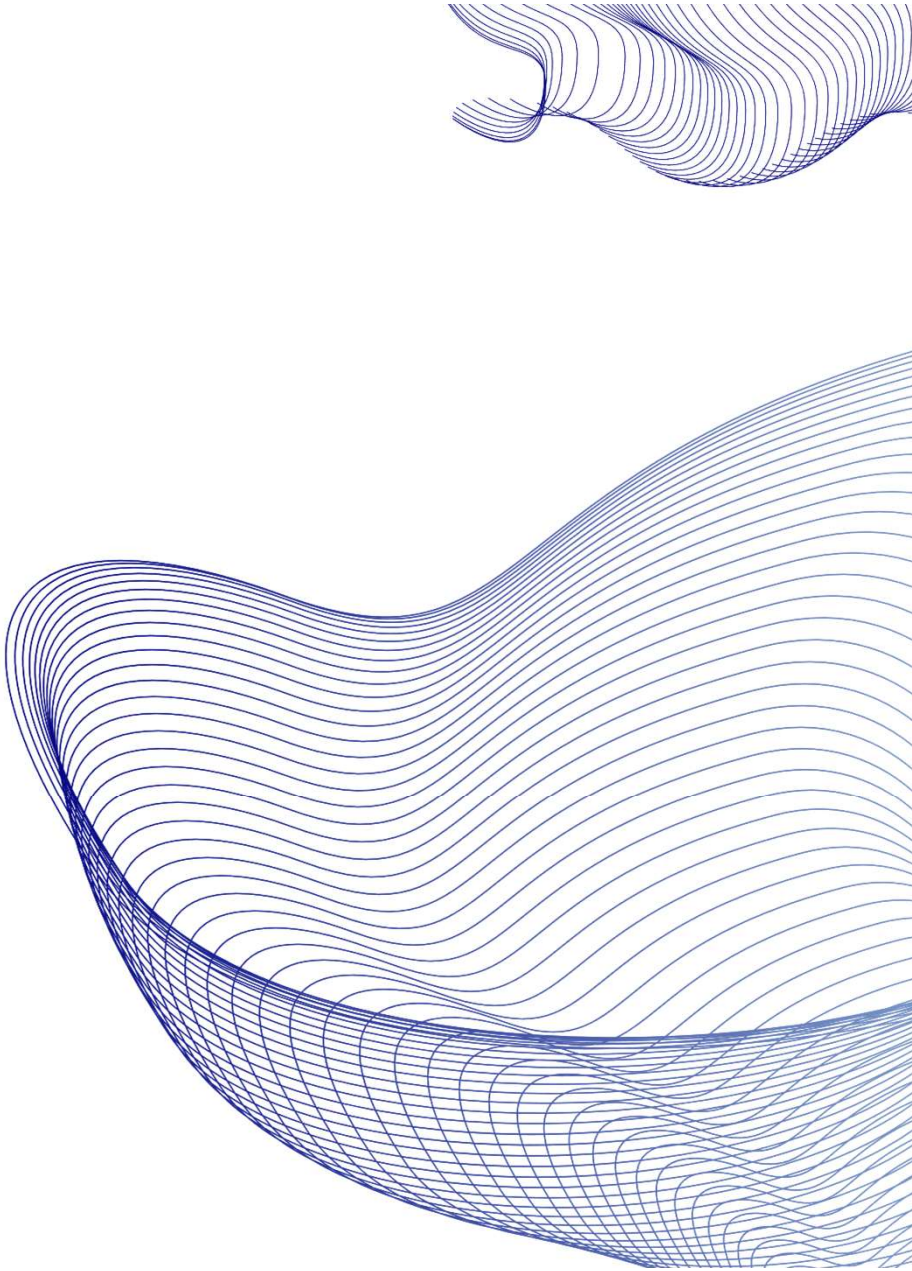
CPCJ

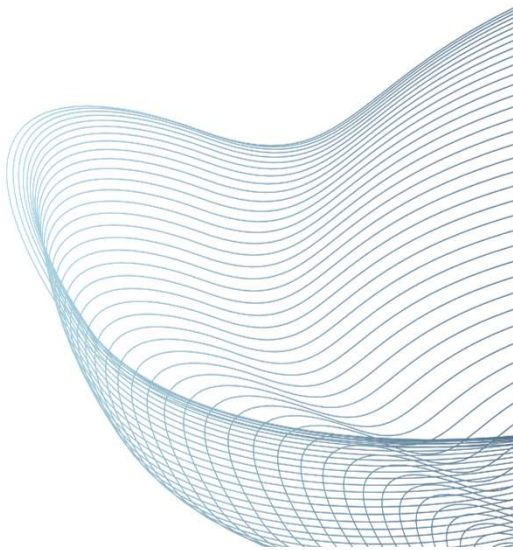
Forecast in a Bank/Finance Sector

Business Data Challenge: CCH Tagetik
Course: Fundamentals of Business Management
Accademic Year: 2023/2024



Our Team

- **C**attari Simona
- **P**oiani Marco
- **C**arella Alessandro
- **J**allow Ebrima



CPCJ

Challenge Presentation

Problem

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.

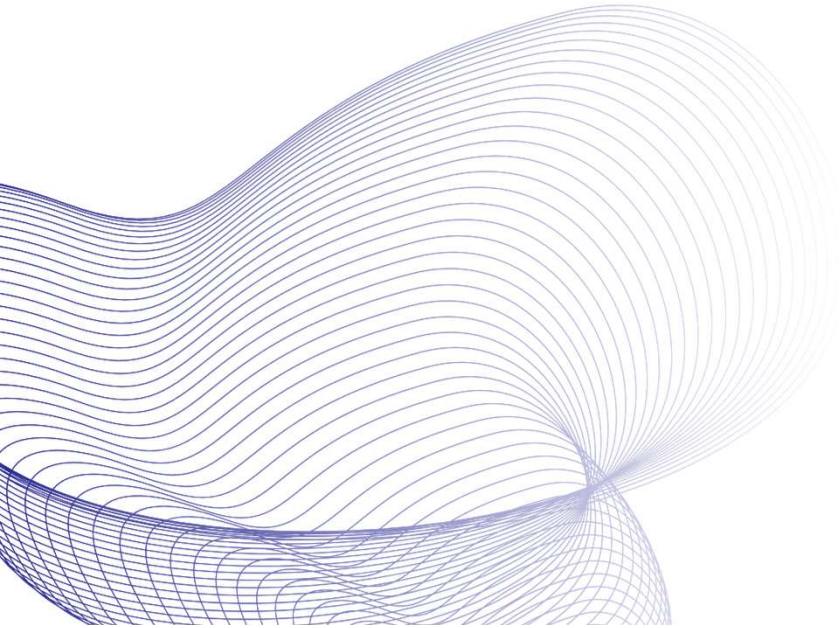
Our Goal

Mock up of a dashboard with KPIs useful for a what-if analysis after the forecast results considering the assumptions changes and the impacts on the EBIT

CPCJ

Tasks

These are the main 4 macro-tasks necessary to solve our problem.



Task 1

Choice of the most significant P&L items as collection drivers

Task 2

Choice of the most significant macro economic variables.

Task 3

Preparation and choice of one or more ML models for the train and forecast phases.

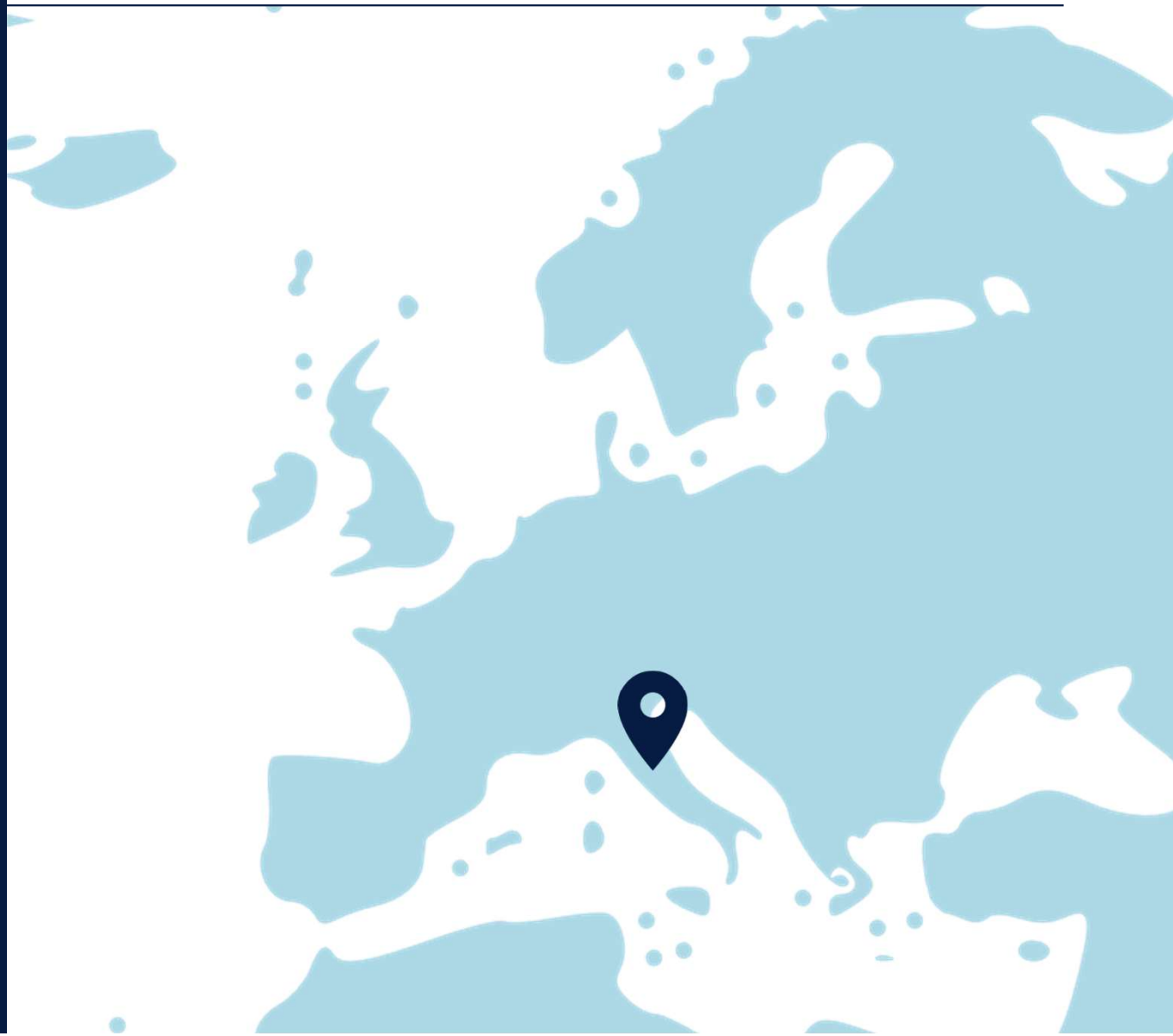
Task 4

Choice of output evaluation metrics.

Target Banks

We built our dataset extracting data from the financial reports of the first 5 banks in Italy by assets. Data spans from 2018 to 2022.

Unicredit
Intesa Sanpaolo
MPS
Bper Banca
Banco BPM

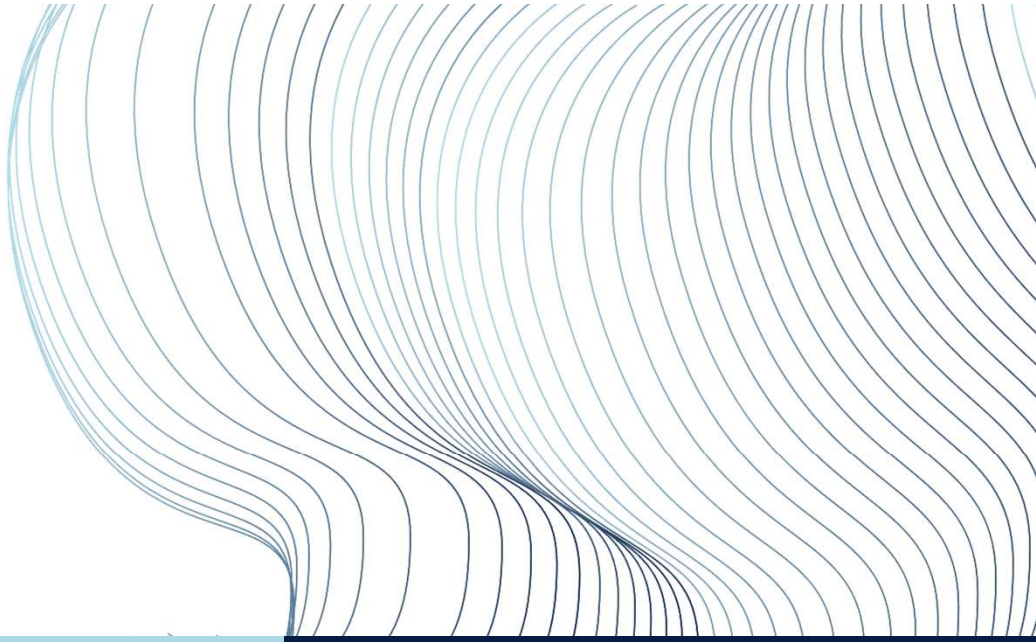


CPCJ

Financial Statements

The selection of the correct items from financial statements has been a determinant for the success of our project.

We decided to extract information from the Income Statement, also known as the P&L statement, because it contains all the necessary elements.



Consolidate P&L statement

Tipically banks are groups, so they must prepare a consolidated Income statement.

The distinction between the income statement and the consolidated version is that the first one includes only the financial results of a single company, the second one incorporates the results of both the parent company and its subsidiaries.

Items breakdown

All the items presented so far are in turn made up of multiple values and variables.

For instance, if we break down item “10”, we realize that it is the result of several elements.

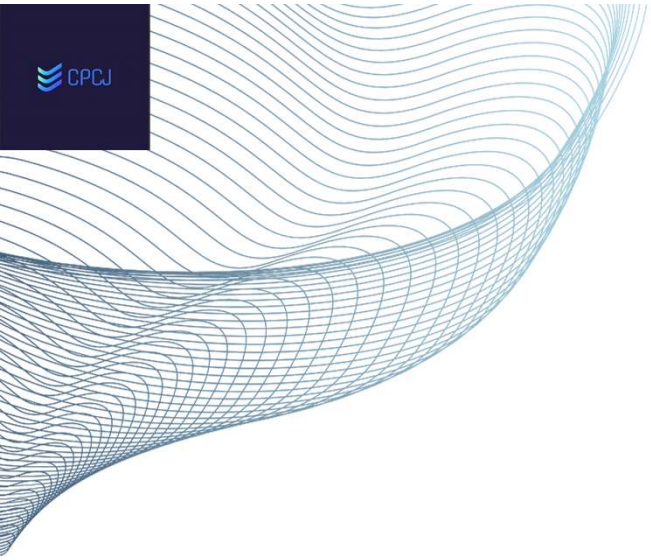
Relevant items

Besides the Net Income and the EBIT we considere other 6 items relevant for our task:

- Net Interest Margin
- Net fees and commisions
- Operating Income
- Net profit from financial activities
- Administrative expenses
- Operating costs

ITEMS	YEAR 2022
10. Interest income and similar revenues	16,339
of which: interest income calculated with the effective interest method	*13,426*

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



KPIs

Generic and useful KPIs for the What-if analysis. For our project, we used those extractable from the Income statement. Exploiting the right KPIs we can measure and analyze the overall performance of a business.

Net Profit Margin

It is a percentage that represents the profitability of a company in relation to its revenue.
It measure of how much profit a company makes for every dollar of revenue it generates.

Operating Expenses Ratio

It is calculated by dividing all operating expenses less depreciation by operating income. A lower operating expense ratio is more desirable for investors because it means that expenses are minimized relative to revenue.

ROA

It measure the overall profitability. It is calculated by dividing net income by average total assets.
We did not use this indicators in our project because the total assets is a number reported on the balance sheet.

Macroeconomic values

Where we have found the values

Official website of banks

Banks are required to maintain a certain level of transparency for their shareholders and the general public.

They have regulatory obligations that require the disclosure of certain financial and macroeconomic data.

Transparency can positively influence the perception of the public and stakeholders.

Financial analysts regularly examine the macroeconomic data of banks to assess their financial strength and ability to manage risks.

Banks that follow these practices can be considered more reliable and professional.

CPCJ

Macroeconomic values

Where we have found the values

Word Bank

An international organization that collects and publishes data reflecting global economic and social conditions.

We used this resource to perform a thorough comparative analysis of macroeconomic variables found in the financial reports of different banks.

This approach provided a detailed understanding of economic dynamics.

Macroeconomic variables used

These are key macroeconomic indicators that play a significant role in shaping the economic landscape.

Gross Domestic Product (GDP): It reflects the economic health of a nation. Its variations influence the demand for financial services and the attitude towards risk. A growing GDP indicates a favorable environment for investments, while a contraction may suggest the opposite.

Unemployment Rate: A key indicator of labor market stability, influences loan demand, consumer confidence, and the real estate sector. High unemployment may lead to reduced loan repayment capacity and a decline in financial services demand.

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: 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.

Exchange 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 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.

CPCJ

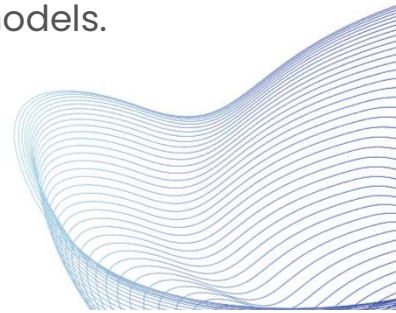
Preparation and choice of ML models

Problem

The task of the project was to build a machine learning model to make predictions based on the retrieved data.
The process of building the model was harder than expected since we could not find a big and therefore reliable amount of data to build the models using the ones my colleagues and I have been working on during our data mining and machine learning courses.

Our Goal

Researching through the internet for models that can be considered reliable with as few data point as the ones we have on our hands I was puzzled by deciding whether to go for a one shot model or linear regression models.



CPCJ

Preparation and choice of ML models

Linear regression

Linear regression is a statistical method used for modeling the relationship between a dependent variable and one or more independent variables by fitting a linear equation to observed data.

The basic idea is to find the best-fitting straight line (or hyperplane, in the case of multiple independent variables) that minimizes the sum of the squared differences between the observed and predicted values

Ridge

Variance of linear regression technique that is an extension of ordinary least squares (OLS) regression

It is used when there is multicollinearity in the data, which means that the independent variables are highly correlated.

The goal is to minimize the sum of squared differences between the observed and predicted values while also penalizing large coefficients.

CPCJ

Preparation and choice of ML models

Lasso

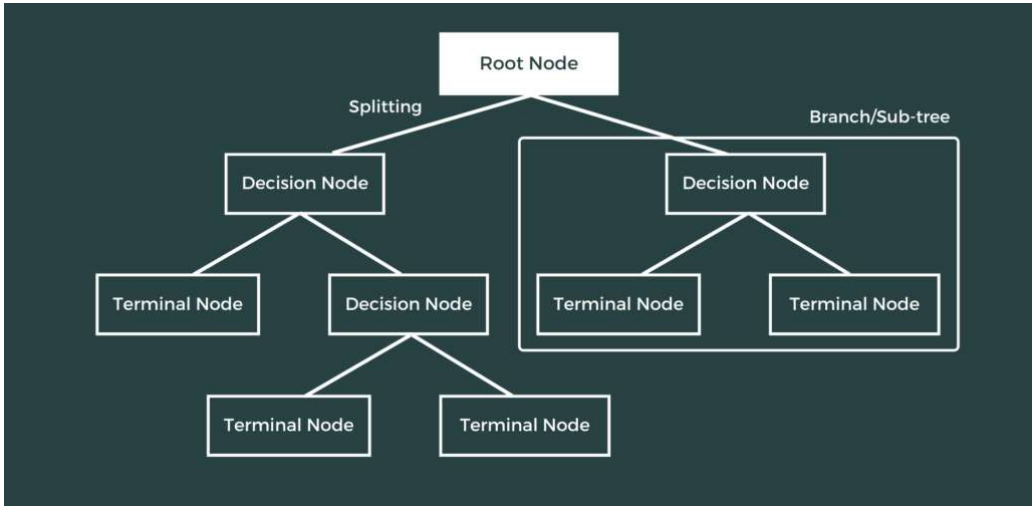
The Lasso (Least Absolute Shrinkage and Selection Operator) is a statistical technique and a regularization method used in linear regression to improve the model's performance and prevent overfitting. It adds a penalty term to the standard linear regression objective function, which helps to shrink the coefficients of some features to zero.

Elastic net

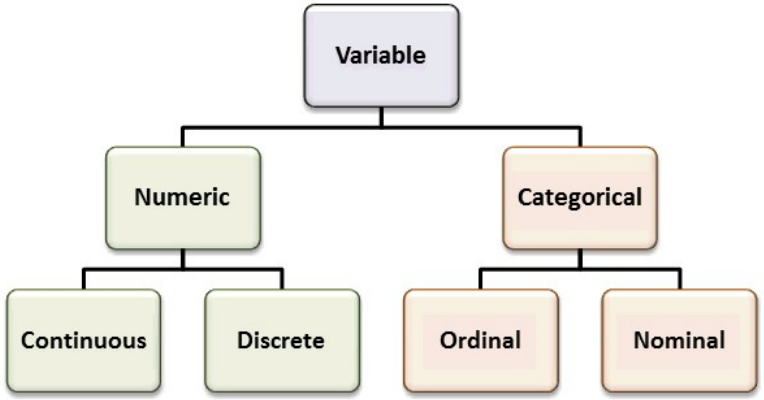
Elastic Net is a linear regression model that combines the L1 (Lasso) and L2 (Ridge) regularization techniques.

Preparation and choice of ML models

Decision tree regressor



Dataset



GDPIndex, UnemploymentRate, PPI, ExchangesRatesPercentage, CPIIndex, CovidStringencyIndex, RealInterestRate

Fixed for a determinate year.

(25, 51)

	InteressiattividaproventiassimilatiDiCuiInteressiAttiviCalcolatiConIlMetodoDellInteresseEffettivo	InteressiPassiviEOneriAssimilati	MargineDiInteresse	CommissioniAttive	...	ExchangesRatesPercentage	CPIIndex	CovidStringencyIndex	RealInterestRate	CostPerEmployee
count	25.000000	25.000000	25.000000	25.000000	...	25.000000	25.000000	25.000000	25.000000	25.000000
mean	6288.933320	5792.892520	-1758.194040	4587.510720	8852.056240	...	0.884000	2.196000	31.580000	0.980000	-0.125240
std	5523.441782	4842.504005	1715.763965	3808.763532	21845.461588	...	0.037417	2.782307	32.757748	0.781025	0.190851
min	990.695000	1358.857000	-5715.000000	1122.437000	812.147000	...	0.850000	-0.100000	0.000000	-0.400000	-0.756000
25%	1975.109000	1863.740000	-3269.000000	1490.354000	1714.284000	...	0.850000	0.600000	0.000000	0.700000	-0.084000
50%	2513.924000	2468.870000	-615.476000	2016.140000	2018.601000	...	0.880000	1.100000	21.990000	1.400000	-0.078000
75%	10486.000000	10565.000000	-460.005000	7993.000000	8105.000000	...	0.890000	1.900000	53.500000	1.600000	-0.071000
max	16339.000000	13426.000000	-192.233000	10853.127000	112087.000000	...	0.950000	8.200000	82.410000	1.600000	0.090000

Choice of output evaluation metrics

Problem

Determining which model is the better was a hard task since the total number of data point is too low to perform a k fold cross validation, so I decided to find multiple metrics and other measures to compare the various models implemented

Mean Squared Error (MSE)

·widely used metric for evaluating the performance of prediction models. When comparing different models, MSE provides a quantitative measure of the average squared differences between the predicted values and the actual values. The lower the MSE, the better the model's predictions align with the true outcomes. MSE is particularly useful in regression analysis, where the goal is to minimize the sum of squared errors.

Choice of output evaluation metrics

Mean Absolute Error (MAE)

valuable metric for assessing the accuracy of a model's predictions. MAE quantifies the average absolute difference between the predicted values and the actual values in a dataset. Unlike other error metrics, MAE is particularly straightforward and easy to interpret, as it represents the average magnitude of errors without considering their direction. A lower MAE indicates better predictive performance, reflecting a closer alignment between the model's predictions and the true outcomes.

R2 score

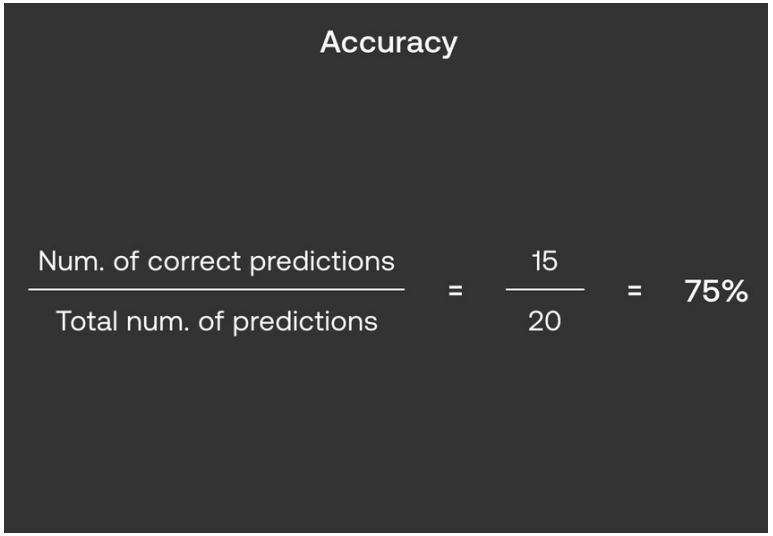
also known as the coefficient of determination, is a metric used to assess the goodness of fit of a regression model. It measures the proportion of the variance in the dependent variable that is predictable from the independent variables. The R2 score ranges from 0 to 1, where a score of 1 indicates a perfect fit, meaning the model explains all the variability in the data. A score of 0 implies that the model does not explain any variability.

Choice of output evaluation metrics

Explained variance score

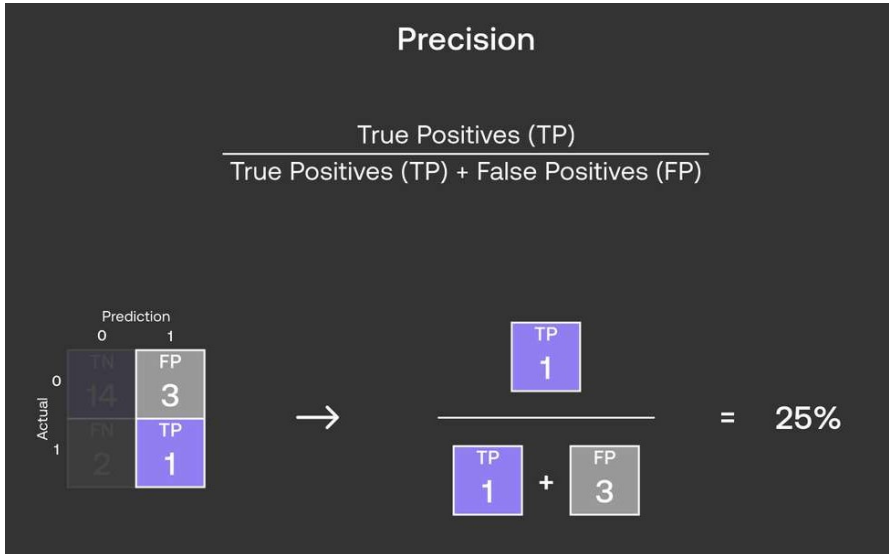
metric used to evaluate the performance of prediction models, particularly in the context of regression analysis. It quantifies the proportion of variance in the target variable that the model can explain. The score ranges from 0 to 1, with 1 indicating a perfect match between the model's predictions and the actual values. While it complements other evaluation metrics, such as mean squared error or R-squared, the explained variance score focuses specifically on the explained variance, offering a concise measure of the model's effectiveness

Accuracy

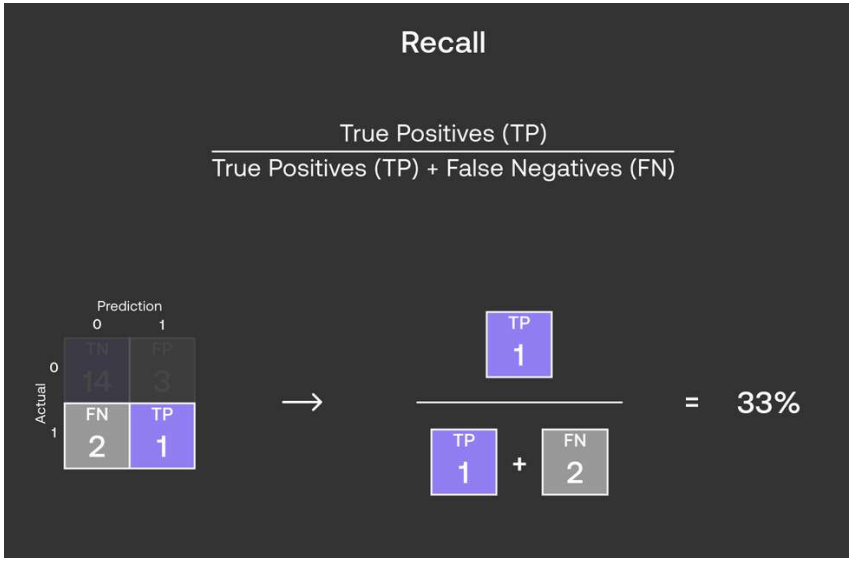


Choice of output evaluation metrics

Precision



Recall



Choice of output evaluation metrics

F1 score

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * 25\% * 33\%}{25\% + 33\%} = 28\%$$

```
Mean Squared Error: [616072.2435259  302263.27150354]
Mean Absolute Error: [657.32054194 352.24824301]
R-squared: [0.92595817 0.95878311]
Explained Variance Score: [0.92829973 0.97233646]
Accuracy: 0.8
Precision: 0.8
Recall: 0.8
F1 Score: 0.8
```

CPCJ

Predictions

To make predictions via the model on the dataset we have to set all the other parameters of the data.

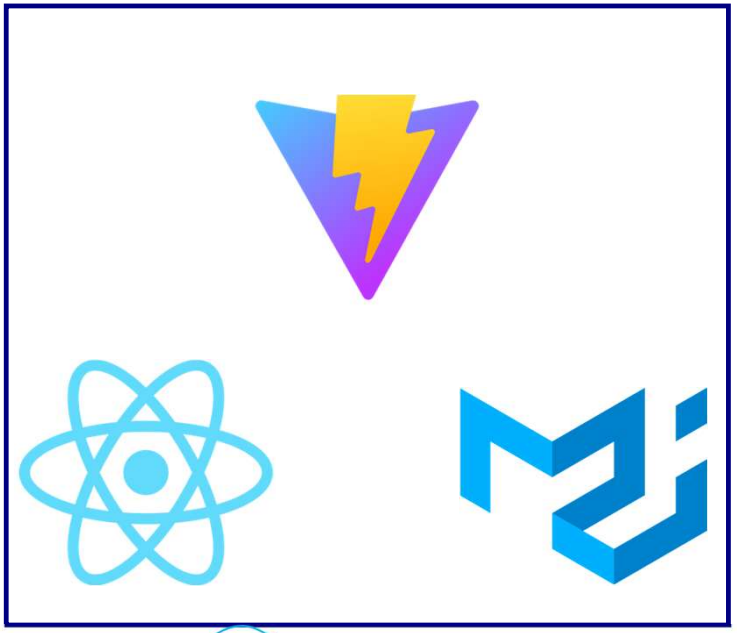
It's better to specify all the known values but, after testing, we decided that having the option to input all the values of the data to predict on the actual developed web application results in a bad user experience.

We decided than to select the most relevant ones and to use a representative value, extracted from the dataset the models are built onto, as a value for the other features when making a prediction.

```
testData = {
    #"Anno": "31/12/2023",
    #"Banca": "UniBank",
    "Interessiattividaproventiassimilati": 15500,
    "DiCuiInteressiAttiviCalcolatiConIlMetodoDellInteresseEffettivo": 12000,
    "InteressiPassiviEOneriAssimilati": -5000,
    "MargineDiInteresse": 9000,
    "CommissioniAttive": 7500,
    "CommissioniPassive": -1200,
    "CommissioniNette": 6000,
    "DividendiEProventiSimili": 400,
    "RisultatoNettoDellattivitaDiNegoziazione": 800,
    "RisultatoNettoDellattivitaDiCopertura": 350,
    "Utili(perdite)DaCessioneORiacquistoDi": 400,
    "AttivitaFinanziarieValutateAlCostoAmmortizzato": 120,
    "AttivitaFinanziarieValutateAlFairValueConImpattoSullaRedditivitaComplessiva": 120,
    "PassivitaFinanziarie": 180,
    "RisultatoNettoDelleAltreAttivitaEPassivitaFinanziarieValutateAlFairValueConImpattoAContoEconomico": 500,
    "AttivitaEPassivitaFinanziarieDesignateAlFairValue": 1000,
    "AltreAttivitaFinanziarieObbligatoriamenteValutateAlFairValue": -550,
    "MargineDiIntermediazione": 18000,
    "RettificheERipreseDiValoreNettePerRischioDiCreditoDi": -1900,
    "AttivitaFinanziarieValutateAlCostoAmmortizzato2": 500,
    "AttivitaFinanziarieValutateAlFairValueConImpattoSullaRedditivitaComplessiva2": 500,
    "UtiliEPerditeDaModificheContrattualiSenzaCancellazioni": -1800,
    "RisultatoNettoDellaGestioneFinanziaria": -50,
    "RisultatoNettoDellaGestioneFinanziariaEAssicurativa": -5,
    "SpeseAmministrative": 16000,
    "SpesePerIlPersonale": 160.5,
    "AltreSpeseAmministrative": -9500,
    "AccantonamentiNettiAiFondiPerRischiEOneri": -5500,
    "ImpegniEGaranzieRilasciate": -4000,
    "AltriAccantonamentiNetti": 30,
    "RettificheERipreseDiValoreNetteSuAttivitaMateriali": 40,
    "RettificheERipreseDiValoreNetteSuAttivitaImmateriali": -10,
    "AltriOnerERoventiDiGestione": -700,
    "CostiOperativi": -500,
    "Utili(perdite)DellePartecipazioni": 550,
    "Utili(perdite)DaCessioneDiInvestimenti": -10000,
    #"Utile(perdita)DellaOperativitaCorrenteAlLordoDelleImposte": 270,
    "ImposteSulRedditoDellesercizioDelloperativitaCorrente": 30,
    #"Utile(perdita)DellaOperativitaCorrenteAlNettoDelleImposte": 7000,
    "Utile(perdita)DiEsercizio": -800,
    "PersonaleDipendente(valoreAssoluto)": 6200,
    "GDPIndex": 6200,
    "UnemploymentRate": 80000,
    "PPI": 3.2,
    "ExchangesRatesPercentage": 7.8,
    "CPIIndex": 24,
    "CovidStringencyIndex": 0.90,
    "RealInterestRate": 8.50,
    "CostPerEmployee": 20.99,
}
```

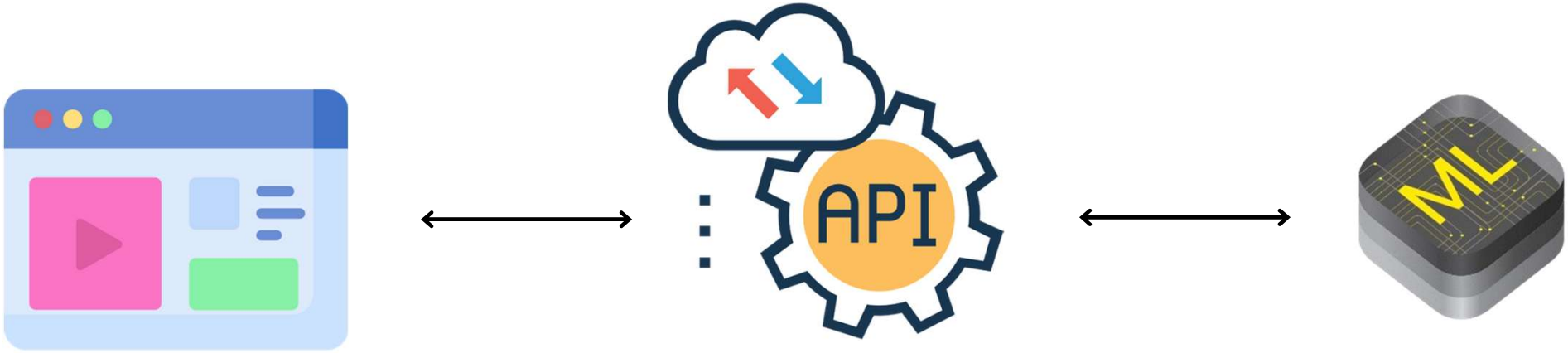
CPCJ

Integration of the systems

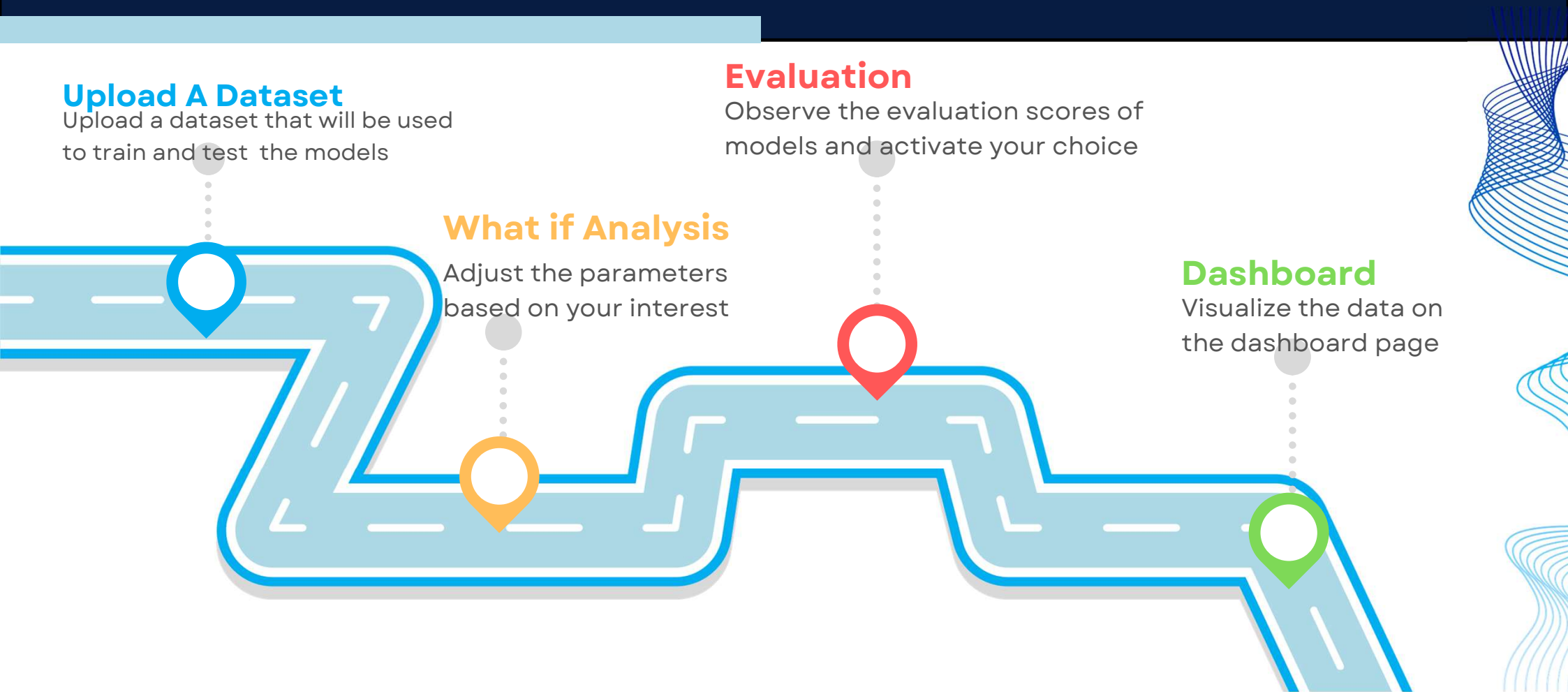


CPCJ

Technical Process



User Flow Illustration



Upload A Dataset

Upload a dataset that will be used to train and test the models

What if Analysis

Adjust the parameters based on your interest

Evaluation

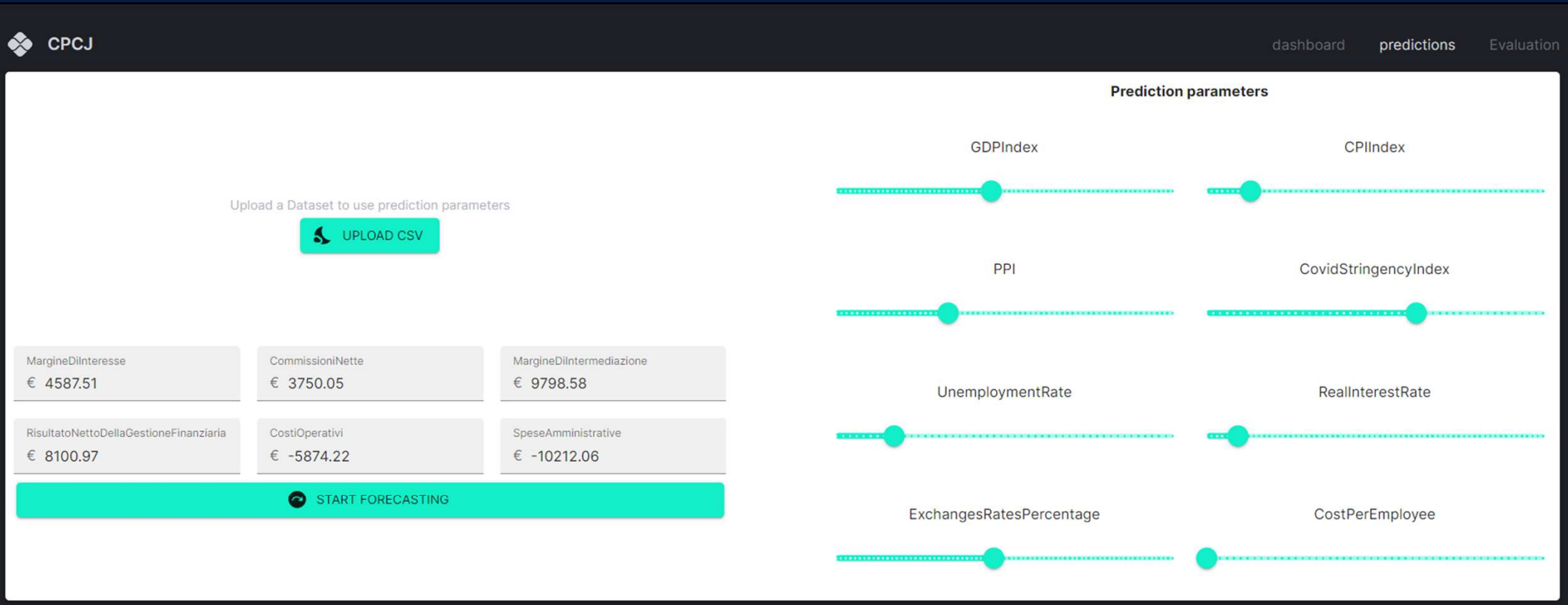
Observe the evaluation scores of models and activate your choice

Dashboard

Visualize the data on the dashboard page

CPCJ

Prediction: What-if Analysis



Forecasting

CPCJ

Evaluation: User Activate a model

CPCJ

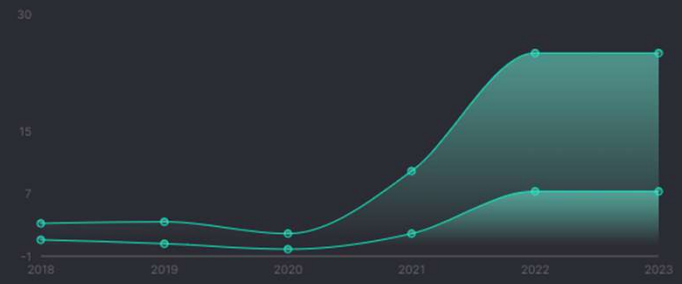
Dashboard

CPCJ

dashboard predictions Evaluation

PPI vs CPI from 2018 - 2023

top line represents PPI, bottom line represents CPI



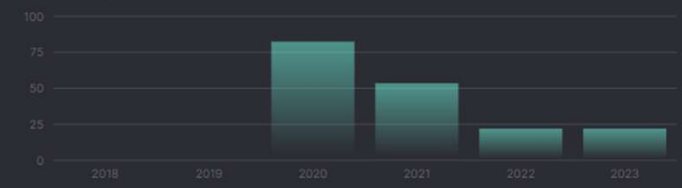
Earning before Income tax and Net Income

top line represents EBIT, bottom line represents Net Income

Strigency Index by Year

graph representing the strigency index from 2018 - 2022

0%



GDP vs Unemployment Rate

+4%

Profit Margin

Percentage

17,45 %

Finance profit of the fiscal year

Profit of the year (millions)

1435

Net Profit - Financ. Activit. (millions)

8220

Exchangre Rate vs Interest Rates

top line represents Interes Rates, bottom line represents Exchange Rate (FX)

+4%

Revenue before Income tax (millions)

Year	Earnings Before Income Tax (EBI...	% Taxed	Revenue
2018	€ 729	5%	€ 692
2019	€ 734	3%	€ 711
2020	€ 748	9%	€ 683
2021	€ 656	0%	€ 656

Employees

Bank	Number of Employees	Cost Per Emp.	Total Cost per Emp.
bank 1	87352	0.04	37777
bank 2	20825	0.08	1648
bank 3	10751	0.34	36500
bank 4	91338	0.07	12504
bank 5	20563	0.08	33584

Operating Expense Ratio

Percentage

71,45 %

Operating Cost (millions)

5874

Revenue (millions)

8220

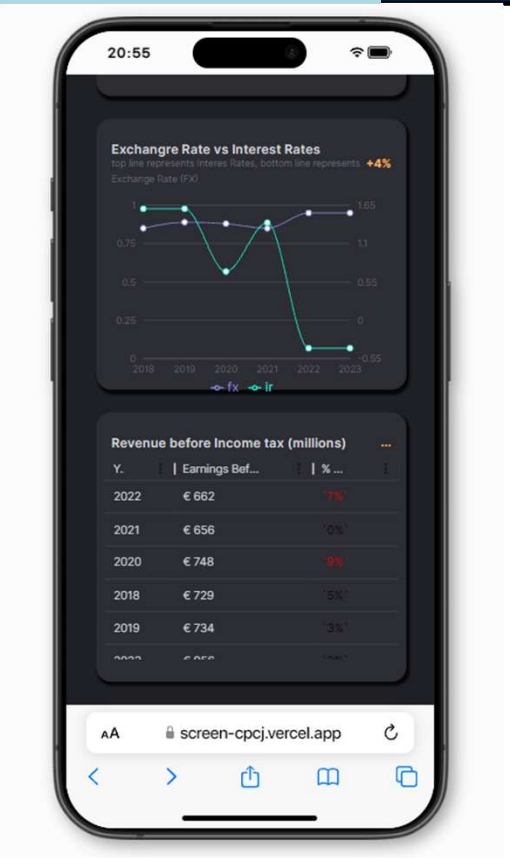
Expected Net Income Compared to Last Year

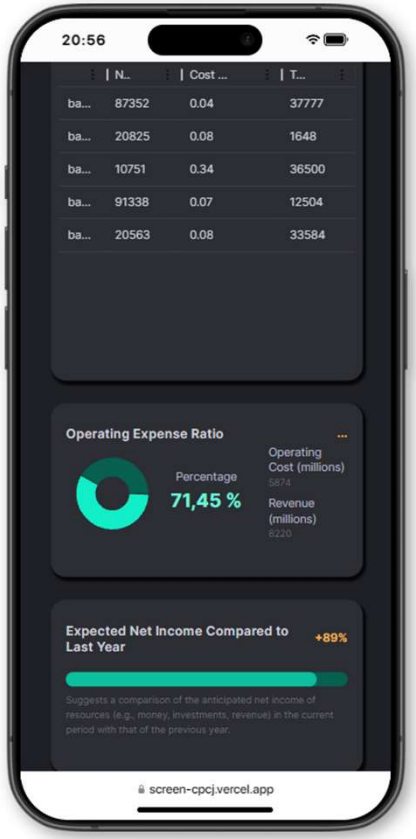
+89%

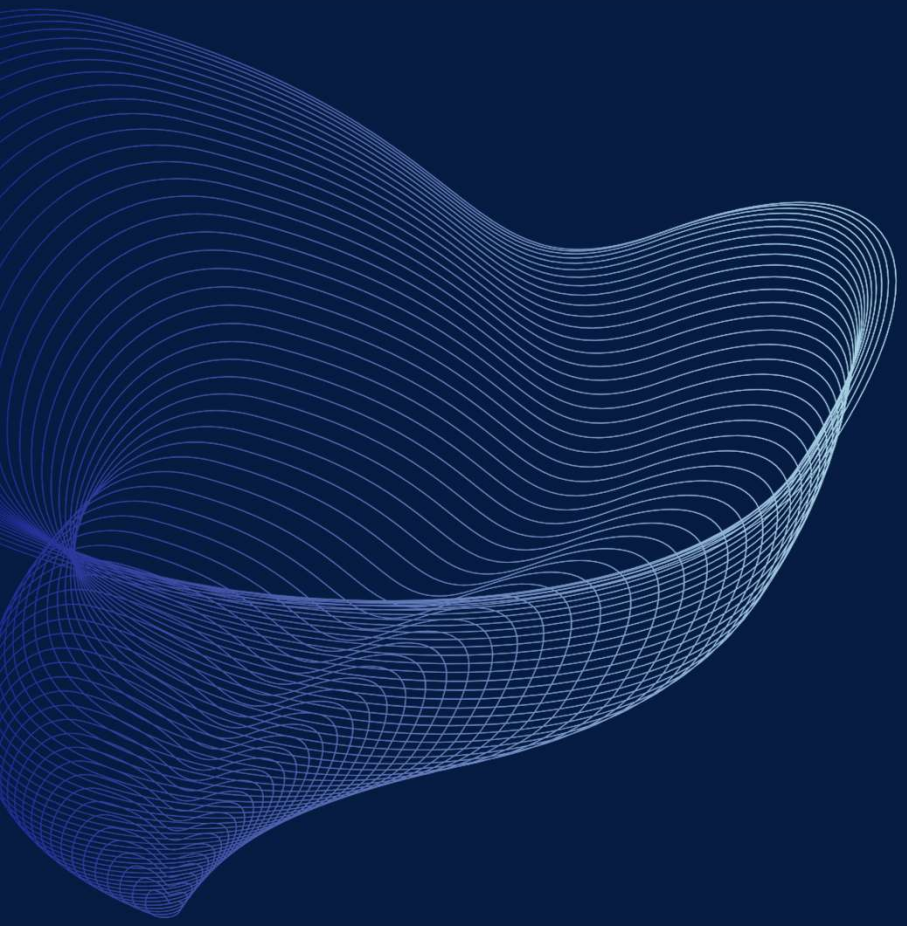
Suggests a comparison of the anticipated net income of resources (e.g., money, investments, revenue) in the current period with that of the previous year.

CPCJ

Responsive Site







Thank you.