Exploring the relationship between Environmental, Social and Governance (ESG) factors and the occurrences of financial crises using machine learning techniques

BSc. Computer Science

Candidate Number: 203068

School of Informatics and Engineering, University of Sussex

Project Supervisor: Dr. Adam Barrett

May 2021

Statement of Originality

This report is submitted as part requirement for the degree of computer science at the University of Sussex. It is the product of my own labour except where indicated in the text. The report may be freely copied and distributed provided the source is acknowledged.

Candidate: 203068

Signed: D. Simons

Date: 18th May 2021

Acknowledgements

I would like to thank my supervisor Dr Adam Barrett for supporting me throughout the production of my final project and dissertation.

I would also like to thank Rachael for providing invaluable insight into the econometrics industry and letting me use datasets from her company.

I would also like to thank Dr Antoniades for allowing me to use his Adjusted Multidimensional Poverty Framework dataset that has proved invaluable for training and testing my machine learning models.

Summary

Financial crises occur in many different forms, locations, timeframes and on a huge range of scales. Predicting and preparing for them is a full-time task for many economists and governmental organisations. Many different institutions have tried to formulate ways to reduce the impact of crises as well as predict warning signs that may hint one is approaching. This research project will attempt to predict sovereign banking crisis occurrence by taking an unconventional approach.

The initial primary goal of this research project was to implement a machine learning model that could predict the occurrence of various financial crisis events through the unconventional approach of using Environmental, Social and Governance (ESG) data to train machine learning models. This has never been successfully done.

Ultimately, it proved challenging, and it is likely that is somewhat impossible to fully achieve. However, conclusions can be draw from the various experiments that could provide useful for future research into the area. For example, many results concluded from the following report could help when looking at key ESG features that could increase or decrease the chances of financial distress.

Table of Contents

1. Introduction

1.1 Project Overview

Environmental, Social and Governance (ESG) factors are becoming increasingly important when it comes to analysing countries and their current levels of development, above and beyond conventional economic measures. However, there is little research exploring the link between these extra-financial indicators and traditional financial events such as financial crises (Crifo, Diaye and Oueghlissi, 2015) (Broadstock *et al.*, 2021).

The overarching objective behind this project is investigating the link, if any, between a country's ESG indicators and the chance of a financial crisis occurring in said country using advanced ML techniques. For the purposes of this report Environmental, Social and Governance indicators will be measured using various metrics. Environmental development will be measured using indicators such as amount of protected forestry area and levels of greenhouse gas emissions. Social development will be measured using levels of access to basic needs (e.g., access to safe drinking water), levels of healthcare (e.g., child mortality rate) and levels of education (e.g., percentage of children in primary education). Levels of Governance development will be assessed based on a variety of factors including levels of foreign direct investment (FDI) as well as various measures of debt.

Investigations will cover three main types of financial crises: currency, sovereign debt, and banking crises. For this report these crises will follow the definitions outlined in Antoniades et al.'s research (Antoniades, Widiarto and Antonarakis, 2019). They define a currency crisis 'as a nominal depreciation of the country's currency vis-à-vis the U.S. dollar of at least 30 percent that is also at least 10 percentage points higher than the rate of depreciation in the year before'. Banking crises considered in their research are restricted to only systemic ones which occur 'when two conditions are met: there are significant signs of financial distress in the banking system e.g., bank runs and bank liquidations and significant banking policy interventions in response to banking losses e.g., deposit freezes'. Lastly, sovereign debt crises which they define as events that 'include episodes of sovereign default or debt restructuring'.

1.2 Literature Review

Existing studies have proved a link between traditional econometric measures and the occurrence of crises. For example, a strong link has been proved with credit growth and the yield curve using ML techniques through the implementation of various models including random forest and support vector machines (Bluwstein *et al.*, 2020). Credit growth is defined as the rising of the amount of money a country is borrowing. The yield curve is defined as 'a line that plots interest rates of bonds having equal credit quality but differing maturity dates' (Chen, 2020). Methodologies utilised by the Bank of England study may prove helpful in this project however their approach was focused on global financial events whereas this research intends to focus on domestic financial crisis events.

The study "sovereign bond spreads and extra-financial performance" (Margaretic and Pouget, 2018) investigates whether emerging countries with good ESG performance have a lower risk of default and therefore a lower cost of debt (both of which mean they are less likely to experience a financial crisis). For Margaretic et al.'s research they measure environmental performance by 'how well countries manage their natural resources'. Social performance in their research 'measures the countries' human development'. Finally, governance performance is measured using the World Bank's World Governance Index (World Bank, 2020). To measure environmental performance, they use Yale's Environmental Performance Index (EPI) which is a data driven summary of the state of a nation's sustainability using 32 performance indicators. These indicators cover a range of measures of environmental health including factors that track levels of environmental damage that impact human health as well as levels of biodiversity. Countries are given an EPI score from 0 to 100 (100 being the best).

They also use the JP Morgan's Emerging Mark Bond Index Global (*Index Suite*, 2020) to calculate the total return performance of bonds issued by emerging market countries. This research concluded that emerging markets with good ESG performance are associated with a lower cost of debt particularly when focusing on environmental and social performance. This would therefore indicate a link between ESG indicators and the chance of financial crisis occurring in any given nation. However, it must be noted that this study does not make that direct link (which this research does hope to achieve). It also must be noted that it only focuses on emerging markets whereas this research project intends to focus on all markets across the globe and therefore more data will be needed for this investigation.

This report is drawing upon much of the data used by Antoniades et al. . (Antoniades, Widiarto and Antonarakis, 2019) which is based upon measures assessing whether the United Nations (UN) has attained their Sustainable Development Goal (SDG) of eradicating poverty. The SDGs were set up in 2015 and consist of 17 goals outlined by the UN that they aim to be achieved from 2020 till 2030. The UN describes them as a "blueprint to achieve a better and more sustainable future for all" (*SDG Indicators — SDG Indicators*, 2017). The SDGs are monitored by 169 different individual target goals that make up the 17 broader goals. Eradicating poverty by 2030 is the first SDG goal. For the purposes of Antoniades et al's research they created an adjusted Multidimensional Poverty Framework (MPF) that included 15 indicators that incorporated key poverty aspects related to income, basic needs, health education and the environment *(See figure 1)*. A regression-based econometric model was then utilised to enable them to examine what impact financial crisis events had on the MPF indicators. The model covered crises in 150 countries during a 35-year period (1980-2015). Their results highlighted a clear link between the occurrence of financial crisis events and a decline in environmental and social indicators. The study concluded that financial crises could be linked to an approximately 10% increase in extreme poverty in low-income countries. They also find that the crises are associated with an "average decrease of government spending in education by 17.72% in low-income countries". Antoniades et al.'s research is key to informing the work carried out in this project since it is essentially carrying out the opposite of what this project is trying to achieve as they are trying to prove a link between financial crises impact on ESG indicators and this project hopes to prove a link between ESG indicators and when a financial crisis may occur in any given country. Therefore, their Adjusted MPF is a great source for the ESG and financial crisis data, especially given they differentiate between the three types of financial crises being differentiated between in this project.

Sustainability Science

Table 2 Adjusted Multidimensional Poverty Framework

Dimensions of poverty	Indicators	Associated SDG goals	Dataset sources	Literature used in the modelling of each indicator
Income	Poverty headcount at $1.90 a day	**1.1**	WDI	Kaasa (2003), Sen (1976)
	Poverty gap at $1.90 a day	1.1	WDI	Kaasa (2003)
Basic needs	Access to safe drinking water	1.4, 3.9 and **6.1**	EPI-Yale	Dube and January (2012), Wrisdale et al. (2017), Alexander et al. (2013)
	Access to basic sanitation	1.4, 3.9 and **6.2**	EPI-Yale	Streeten (1979), Wrisdale et al. (2017), Alexander et al. (2013)
	Access to electricity	1 and **7.1**	WDI	Kemmler (2007), Poloamina and Umoh (2013), Borenstein (2012)
Health	Infant mortality rate (per 1000 live birth)	**3.2**	WDI	Pelletier et al. (1995), Rice et al. (2000), Rutstein (2000)
	Maternal mortality ratio (per 100,000 live births)	**3.1**	WDI	DiOrio and Crivelli-Kovach (2014), WHO (2019), Slocumb and Kunitz (1977)
	Particulate emission damage (% of GNI)	11.6 and 13.2	WDI	Afzal et al. (2014), Zhang and Jiang (2018), Zhou and Levy (2007)
Education	Children out of school (% of primary school age)	1 and 4.1	WDI	Burke and Beegle (2004), Okumu et al. (2008), Siddiqui and Iram (2007)
	Government education expenditure (current US$)	4	WDI	Busemeyer (2007), Chakrabarti and Joglekar (2006), Imana (2017)
Environment	Agricultural land (1000 ha)	**2.4** and 13	CCI	Allahyari and Koundinya (2013)
	Net forest land CO2 emissions/removals (terragrams)	15.2	WDI	Achard et al. (2004), Buys et al. (2017)
	Carbon dioxide damage (current US$)	9.4 and 13	WDI	Loria (2018), Ghouali et al. (2015), Liu et al. (2013), (Al-mulali 2012)
	Forest rents (% of GDP)	15.2 and 12.2	WDI	Imai et al. (2018), Angelsen and Wunder (2003)
	Terrestrial protected areas (global biome weights)	**15.4**	EPI-Yale	Schulze et al. (2018)

Bold in the column 'Associated SDG Goals' indicates an official SDG indicator

Figure 1: Antoniades et al.'s Adjusted MPF

In a more alternative approach, also by the Bank of England, ML techniques in the form of non-linear text analysis on newspaper articles has also been proved to be effective at predicting macroeconomic indicators that can, in turn, be used to predict upcoming financial crises (Kalamara *et al.*, 2020). As stated in this study this approach can be used to 'enhance statistical economic forecasts of growth, inflation and unemployment' using supervised ML techniques. However, techniques utilised in their work have not proved to be effective in this project due to their algorithms implementing natural language orientated techniques which is different to the ML approach taken in this project which will primarily involve numerical data rather than text-based analysis.

Pacca et al. have conducted research regarding financial crises and their effect on the environment in the paper 'the effect of financial crises on air pollutant emissions: 'An assessment of the short vs. medium-term effects'(Pacca *et al.*, 2020). In effect, this paper conducts the reverse of what part of this project is investigating – the relationship between a country's environmental development indicators and financial crises. Despite the opposing research objectives, the results are still relevant and must be considered. Pacca et al. found, in the short-term, financial crises decrease emissions. However, medium-term results indicated that the effects of financial crises were insignificant on emissions, and, in some nations, it led to an increase in emissions compared with pre-crisis levels. In this project results specifically relating to environmental indicators and financial crises are compared with the findings from this study to see if there is any relationship. Their paper will make a good comparison since they are using the same data on financial crises and similar data from the world bank regarding the environment.

Another key piece of research that has influenced the direction of this investigation is Kate Raworth's *'Doughnut Economics'* which is a visual model depicting sustainable development as shown in figure 2 (Rockström *et al.*, 2009). This diagram and her research highlight the importance of remaining in the 'safe and just space for humanity' by ensuring that our ecological ceiling is not abused or 'overshot'. A crucial way to help ensure this does not occur is by highlighting the importance of investing in ESG areas to governments to show that these areas are not only important for the preservation of human civilisation but can also have shorter term benefits such as possibly preventing crises – a key political issue that politicians could use to their advantage.

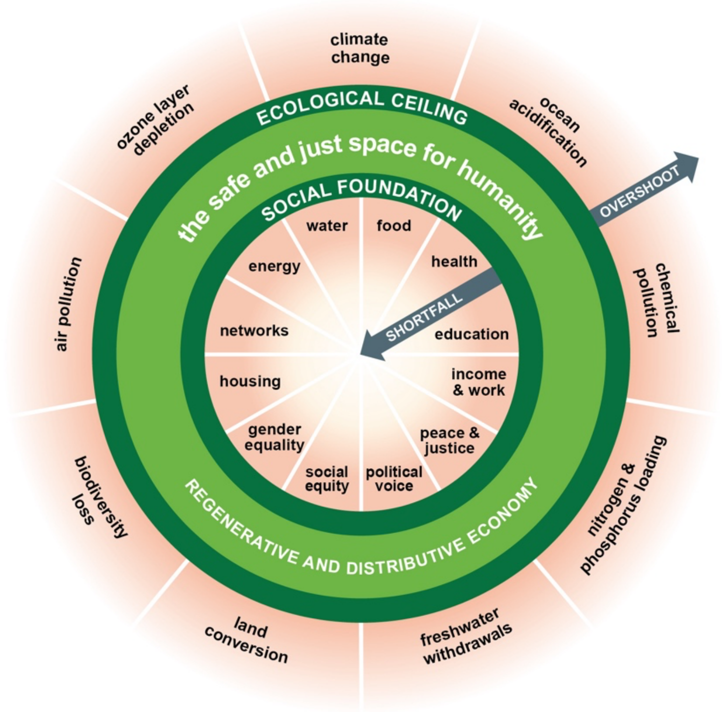

Figure 2: Kate Raworth's Doughnut Economics Visualised

1.3 Motivation

The motivation behind this project is twofold; to highlight the importance of extra-financial indicators, and to perhaps challenge the current thinking that traditional economic metrics are the best way to indicate development. A desired outcome of this research would be to highlight that many measures of less tangible development can only be seen when looking at extra-financial indicators that incorporate a broader overview of a country's more general progress. Proving strong links between ESG factors and financial crises could help inform policy decisions made by governments around the world. Specifically, it could add increased pressure on nations to focus on investing in areas such as environmental and social development since there would be proven economic incentive to better a country, whilst also investing in areas that benefit individuals in a more meaningful way than macro-economic oriented policy does.

1.4 Problem Specification

1.4.1 Research Hypotheses

The main hypothesis for this research project is that ESG indicators have an influence on the occurrence of financial crises. Through the implementation and/or the combination of various ML models this will be attempted to be demonstrated. Since there is little evidence or existing research of this hypothesis it is not clear if this will be possible.

Another hypothesis that will be explored is that traditional metrics of economic good health, for example Sovereign Credit Default Swaps (CDS) can be used to predict financial crises. A sovereign CDS is a financial swap agreement that the seller of the CDS will compensate the buyer in the event of debt default (in this case if a sovereign nation defaults on its debt) or other credit event ('Credit default swap', 2021). This was established by Bluwstein et al (Bluwstein *et al.*, 2020) but will be reproduced in this project to confirm their findings and to establish a framework for crisis prediction to be used when incorporating ESG factors into the prediction model.

The final extension hypothesis will be to establish a link between specific areas of ESG development and the occurrence of financial crises. For example, there is possibly a link between governance and social indicators (such as rates of gender violence) and the occurrence of currency crises.

1.4.2 Website Requirements Analysis

1) Site will be accessible to the public
2) Site will have simple navigation bar
3) Site will display results in form of tables and graphs
4) Site will display final report

2. Professional Considerations

This project has met all the ethical considerations outlined on the ethical compliance form as no participating parties were required to complete any part of the project. The British Computer Society's (BCS) code of conduct (*BCS Code of Conduct | BCS*, 2020) has been taken into consideration wherever it applied.

This project has drawn on data from various sources and throughout section 1b of the BCS code of conduct has been upheld as "due regard for the legitimate rights of Third Parties" has been applied wherever applicable. This has been demonstrated in the form of appropriate referencing to data sources and any external resources. Section 1c of the code – "promote equal access to the benefits of IT and seek to promote the inclusion of all sectors in society wherever opportunities arise" – has also been upheld since the results of this report have been made available for public reference.

Section 2e of the BCS, “respect and value alternative viewpoint and, seek, accept and offer honest criticisms of work” has been upheld throughout the duration of this project. This has been ensured by conducting a thorough literature review and diligent referencing where appropriate. Furthermore, honest criticism has been sought through meetings with persons in industry and during regular supervisor meetings. Section 2f – “avoid injuring others, their property, reputation, or employment by false or malicious or negligent action or inaction” – has also been ensured throughout this project. Section 3c: “accept professional responsibility for your work and for the work of colleagues who are defined in a given context as working under your supervision,” has also been adhered to. Section 4a under the ‘Duty to the Profession heading’ – “accept your personal duty to uphold the reputation of the profession and not take any action which could bring the profession into disrepute” has been closely followed and not violated.

3. Methodology

As well as attempting to demonstrate a general correlation between ESG indicators and financial crises, more specific avenues may be explored. For example, the relationship between governance indicators and the chance of sovereign debt and currency crises. Another possible area of interest that could be investigated is the relationship between social indicators (e.g., rates of violence/gender equality) and the chance of currency crises.

3.1 Overview of ML Models

After all the required research and data analysis was completed the implementation of machine learning models began. This started by implementing an ML model using Scikit-Learn’s logistic regression model. After this, alternative models were experimented with including the K-Nearest-Neighbours model as well as a Gaussian Naïve-Bayes classifier. The decision was taken to not attempt to implement a neural network model as the dataset being used here was not large enough for a neural network to be sufficiently trained on and would therefore most likely yield worse accuracy and precision than alternative models. Feature selection and feature pre-processing was also implemented to engineer the data to suit the models in the correct way.

3.2 Overview of Data Pre-Processing

The most important pre-processing utilised was the Synthetic Minority Oversampling Technique (SMOTE) that was implemented using Imbalanced Learn’s SMOTE library (*ADASYN — imbalanced-learn 0.3.0.dev0 documentation*, 2020). Imbalanced Learn’s library is based on the 2002 research by Nitesh Chawla et al. (Chawla *et al.*, 2002). This was used to artificially increase the size of the 1 class to synthetically create a larger set of instances of the 1 class to create a more balanced dataset. This library includes four different types of SMOTE: plain, borderline1, borderline2 and Support Vector Machine (SVM).

Generally, SMOTE works by identifying examples in the vicinity of the feature of the feature space. It then creates a line between two examples in the feature space. An example datapoint is then selected from across this line and this datapoint is then used as an artificial datapoint. Specifically, as described in the book *Imbalanced Learning: Foundations, Algorithms, and Applications*' (He and Ma, 2013) 'SMOTE first selects a minority class instance at random and finds it k nearest minority class neighbors. The synthetic instance is then created by choosing one of the k nearest neighbors at random and connecting a and b to form a line segment in the feature space. The synthetic instances are generated as a convex combination of the two chosen instances a and b'.

As mentioned above, there are four different SMOTE techniques used in experiments for this report. The borderline1 and 2 techniques work by identifying instances of the minority classes that are mis-classified. It then oversamples the misclassified instances to increase the resolution of problematic, misclassified instances.

SVM SMOTE works by using an SVM classifier to find support vectors and create samples based on these vectors. The amount of support vectors to be considered can be defined in the hyperparameters of this method.

In addition to SMOTE Adaptive Synthetic Sampling approach (ADASYN) was also experimented with as an alternative oversampling approach (*ADASYN — imbalanced-learn 0.3.0.dev0 documentation*, 2020) (He *et al.*, 2008). This technique is like SMOTE but with small variations. Instead of generating artificial points that are linearly related to real datapoints, as is done in SMOTE, ADASYN works by identifying sample points and creating random small artificial values around these points. This makes the way ADASYN generates artificial datapoints slightly more realistic than SMOTE for real-world applications (Bhattacharyya, 2018).

Other pre-processing techniques that were utilised were standardisation and normalisation. Normalisation is the process of scaling each individual datapoint to have a 'unit norm' (*sklearn.preprocessing.normalize — scikit-learn 0.24.1 documentation*, 2020). For this investigation 'l2' normalisation was used. This works by applying the square root of the sum of all the squared values being normalised. Applying this function on the dataset both smooths it and reduces rotational invariance (Dorpe, 2018).

Scikit Learn's scaling method works by standardising data along an axis and scaling all values between 1 and 0. This has the benefit of making the dataset fit a Gaussian distribution (which has benefits, particularly for Naïve Bayes Classification). It also creates data with zero mean and unit variance (unit variance is every value in the dataset divided by the standard deviation of the dataset) which is beneficial to all models that will be used (*6.3. Preprocessing data — scikit-learn 0.24.2 documentation*, 2020).

The final, pre-processing technique utilised was a simple feature selection technique which removed features with a large quantity of missing values. Firstly, features with 2000 or more missing values were removed. This reduced the number of features from 162 to 134. Other values were experimented with ranging from removing features with 4000 or more missing values. However, ultimately, the pre-processing removing features with 2000 or more missing values was determined to be the most effective.

3.3 The Datasets

A necessary requirement for an effective machine learning algorithm is good quality datasets. This research project primarily uses Antoniades et al.'s data (Antoniades, Widiarto and Antonarakis, 2019) in the form of their *Adjusted Multidimensional Poverty Framework*. As mentioned in the introduction, this dataset covers a wide range of ESG variables (including health, education, income basic needs and the environment) across 150 different countries dating from 1980-2015. This dataset was chosen as it differentiates between the three different types of financial crises being evaluated for this research project. It also covers a suitably wide range of ESG indicators from specific social indicators (e.g., nutrition, child mortality and years of schooling) to comprehensive environmental indicators (e.g., carbon dioxide damage and areas of protected forest land).

It must be noted that incorporated in the MPF comes from various other third parties including Yale and the World Bank (*World Development Indicators | DataBank*, 2021) (*Welcome | Environmental Performance Index*, 2021). The bulk of the financial data incorporated in the MPF is taken from Laeven and Valencia's paper *'Systemic Banking Crises Revisited'* (Laeven and Valencia, 2018). They compiled data on system banking crises by analysing the intensity of the policy response from governments to label the crises within the correct timeframe. The database of financial events used in the Adjusted MPF is an enhanced version of the "comprehensive global database on system banking crises' developed by Laeven and Valencia in both their 2008 and 2013 papers. This data on financial events suits this investigation since their use of policy response to data crises significantly reduces the use of subjective criteria when dating these events. As pointed out in Laeven and Valencia's paper this gives their database advantages over others on financial crisis events e.g., Caprio and Klingebeil's (Caprio and Klingebiel, 1999).

3.2 Implementation of ML Models

The approach taken by this project was to start with a relatively straightforward ML model, run experiments with that and then experiment with models that take a different approach. This started with the implementation of a Logistic Regression model. After that a K-Nearest Neighbours model was implemented. Finally, a Gaussian Naïve Bayes Classifier was implemented. Throughout experimentation various parameters were adjusted and tweaked according to the models being used. A large variety of pre-processing was also constantly being adapted to ensure the data suited the model being experimented with at the time.

3.2.1 Logistic Regression Model Implementation

The decision to begin experiments using Logistic Regression was taken because it is a well-established and common ('Commonly Used Machine Learning Algorithms | Data Science', 2017) ML model that suited the classification requirements here since logistic regression provides probabilities on binary classification problems and is also capable of being trained on both continuous and discrete data. These features of Logistic Regression made it favourable over similar models such as Linear Regression (which would not be suitable for this binary classification problem since Linear Regression predicts continuous values).

As opposed to Linear Regression (which makes predictions based on a straight line) Logistic Regression fits datapoints to a sigmoid (S-shaped) logistic function. This curve ranges from 0 to 1 and this curve makes a prediction based upon the weight of prediction. It is ideal for the Adjusted MPF dataset since this contains both discrete and continuous data. Unlike Linear Regression, which fits a line using the *least squared method* (which finds the line that minimises the sum of the squares of the Residuals), Logistic Regression does not have residuals that can be calculated, which therefore means you cannot calculate Residuals or Residuals squared. Instead, Logistic Regression utilises Maximum Likelihood Estimation to fit the line.

The maths that underpins the Sckit-Learn Logistic Regression function being used in this investigation comes from the logistic function shown in equation 1 below.

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}},$$

Equation 1

In this function x_0 is the x value of the midpoint of the sigmoid, L is the curve's maximum value and k is the logistic growth rate (gradient) of the curve (see figure 3) ('Logistic function', 2021).

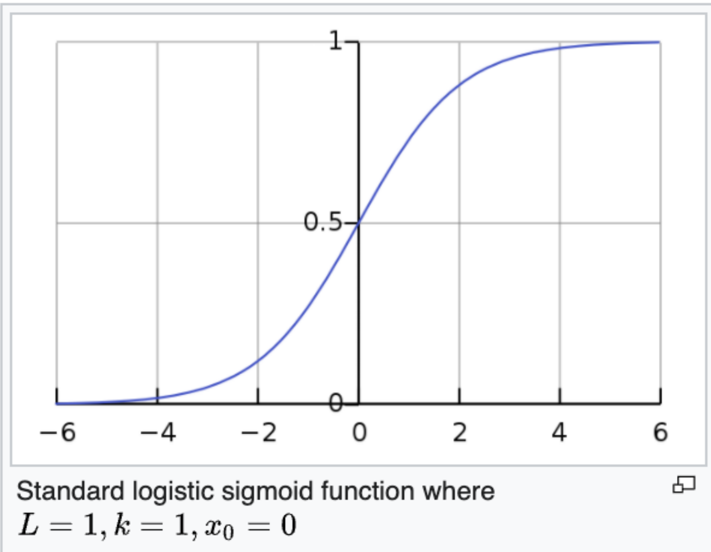

Figure 3: Depiction of Standard Logistic Sigmoid Function

Since probability values (values on the Y-axis) resulting from Logistic Regression models are confined between 0 and 1 the Y-axis is transformed from probability of prediction value to the log of the odds of the prediction value. The resultant Y-axis ranges from – infinity to + infinity. This is done using the *logit function – see equation 2*.

$$\log(odds) = \log\frac{p}{1-p}$$

Equation 2

Coefficients are then presented in terms of the *log(odds)* graph which is a straight-line graph. A candidate best fitting straight line can then be plotted like in Linear Regression. This line has a Y-axis intercept and a slope. However, since this transformation gives the Y-axis a range from – to + infinity it is not possible to use Linear Regression's *least squared method* to find the best fitting line since the Residual values will have the same infinite range of the Y-axis. This is where Maximum Likelihood Estimation (MLE) is utilised.

MLE works by firstly projecting the original data points onto the candidate line generated. This gives each sample a candidate *log(odds)* value. Candidate *log(odds)* can then be transformed to candidate probabilities using the equation 3 formula.

$$p = \frac{e^{\log{(odds)}}}{1 + e^{\log{(odds)}}}$$

Equation 3

This formula is a reordering of the transformation from probability to *log(odds)*. Then to calculate the *log(likelihoods)* the *log(odds)* line is gradually rotated. Each time it is rotated the data is projected onto it. Each time the Y-axis is transformed to probabilities and the log likelihood is calculated. After a few rotations the algorithm eventually finds the line with the Maximum Likelihood which is the line with the best fit.

However, more needs to be done than simply finding the line with the best fit. We also need determine if the line that is fitted represents a useful model which therefore means we need *Residual²* values and *p-values*. This must be achieved without using Residuals (as mentioned above Residual values cannot be determined in Logistic Regression). Currently there is no unified consensus on how to calculate *Residual²* for Logistic Regression (there are more than 10 different ways). For the purposes of this project the model being implemented is Scikit Learn's version (*sklearn.linear_model.LogisticRegression — scikit-learn 0.24.1 documentation*, no date).

3.2.1.1 Logistic Regression Data Pre-Processing

As with all ML models, data pre-processing is essential. For the Logistic Regression model implemented this involved various techniques. It started by identifying the ESG indicators with large number of missing values which would be likely to make the data not very useful. The raw data file had over 150 different indicators spanning the 150 different countries being analysed. Overall, these different indicators the ones with missing data for more than 2/3 of the countries or years within a country were dropped. This reduced the number of columns in the data from 162 in total to 54.

To address the significant imbalance between the 0 and 1 classes SMOTE and ADASYN were experimented with (as discussed at the beginning of the methodology section).

Due to the significant variation in ranges of the data between the different indicators it seemed appropriate to normalise the data before passing it through the model. This was done using Scikit Learn's normalisation (*sklearn.preprocessing.normalize — scikit-learn 0.24.1 documentation*, 2020).

Principle Component Analysis (PCA) was also experimented with for the Logistic Regression model. PCA is a dimensionality reduction technique that aims to reduce the number of variables in the training data set to increase simplicity whilst trying to retain as much information as possible. However, this did not end up being effectively implemented as it suffered from the imbalanced classes (see results section).

PCA reduces the dimensions of a dataset by creating a 'PCA lot' that clusters all the different features in the dataset that have similarities. On this plot axes are ranked according to order of importance. Hyperparameters can then be tuned to specify how many un-important features to remove from the dataset. Although PCA can help to improve model performance it did not help for this investigation - see results section.

3.2.2 K-Nearest Neighbors Model Implementation

The K-Nearest Neigbhbors ML model works is a neighbors based classification technique that does not create its own model but instead has a model represented by the training dataset. It is a type of non-parametric classification method that can be used for both classification and regression ('*k*-nearest neighbors algorithm', 2021). Classification works by calculating a majority vote of the nearest neighbors of each datapoint. For every point analysed it is assigned to the feature class which has the most features within the area that is being defined (this value being *k*). For these experiments, the area is defined by the Euclidean distance which is the distance between a datapoint being looked and another point defined by the algorithm. It calculates the area in which to look around said datapoint by applying the Pythagorean theorem ('Euclidean distance', 2021).

The classifier being used for this project is Scikit Learn's *KNneighborsClassifier* which *"implements learning based on the k nearest neighbors of each query point"* (*1.6. Nearest Neighbors — scikit-learn 0.24.1 documentation*, 2021). For this project the value of *k* was determined by iterating over 40 different possible values of *k*, trying each of these and selecting the one with the value of *k* that resulted in the lowest error rate. This technique was adopted because the optimal value of *k* is extremely dependent on the type and quality of the data being used. It was therefore determined that for the most optimal results the lowest value of k where the minimum error rate was achieved, would be selected.

The value of K that was selected in the end was 12. To visualise this the values of K being tested were plotted on a graph against the mean error rate each model was achieving. As shown in figure 4 the mean error rate levels out at K values of 12 or above. This was therefore chosen by the model since increasing the K value beyond 12 provides not added performance to the model when it comes to decreasing its error rate.

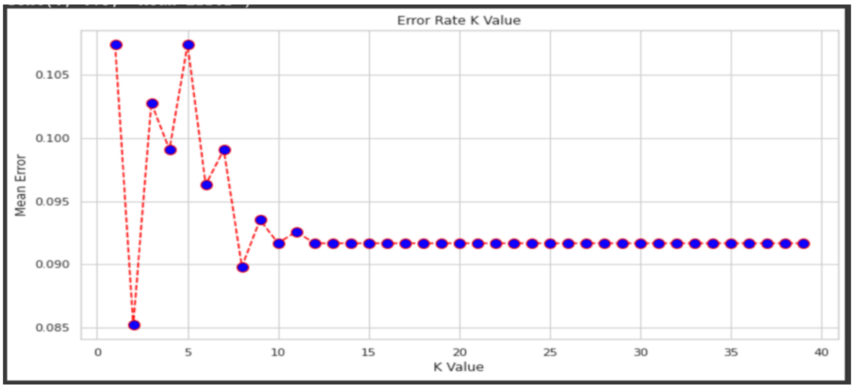

Figure 4: K Value Compared to Mean Error Rate

3.2.2.1 K-Nearest Neighbors Data Pre-Processing

Data pre-process for K-Nearest Neighbors algorithms is extremely important since these algorithms are particularly susceptible to the curse of dimensionality. The curse of dimensionality 'states that in high dimensional spaces distances between nearest and farthest points from query points become almost equal'. This means that if a K-Nearest Neighbors algorithm is trained on a dataset with many dimensions nearest neighbor calculations (for this report these calculations would be the Euclidean distance around the query points) will not be able to distinguish between candidate points. Essentially this means that if data with high dimensionality is used there will be far too much noise around each query point which in turn could lead to meaningless results from the model created (Kouiroukidis and Evangelidis, 2011)

3.2.3 Naïve Bayes Model Implementation

Naïve Bayes models are supervised machine learning models that apply Bayes' Theorem with the '*Naïve'* assumption that all features are independent of each other. The theorem, outlined in figure 5, states the following given a class variable y and dependent feature vector x_1 through x_n (*1.9. Naive Bayes — scikit-learn 0.24.1 documentation*, 2021).

$$P(y \mid x_1, \dots, x_n) = \frac{P(y) P(x_1, \dots, x_n \mid y)}{P(x_1, \dots, x_n)}$$

Using the naive conditional independence assumption that

$$P(x_i | y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n) = P(x_i | y),$$

for all i, this relationship is simplified to

$$P(y \mid x_1, \dots, x_n) = \frac{P(y) \prod_{i=1}^{n} P(x_i \mid y)}{P(x_1, \dots, x_n)}$$

Since $P(x_1, \dots, x_n)$ is constant given the input, we can use the following classification rule:

$$P(y \mid x_1, \dots, x_n) \propto P(y) \prod_{i=1}^{n} P(x_i \mid y)$$

$$\Downarrow$$

$$\hat{y} = \arg\max_y P(y) \prod_{i=1}^{n} P(x_i \mid y),$$

Figure 5: Outline of Naïve Bayes' Theorem

The first line above outlines how to calculate the probability of y occurring given all the feature vectors (x). It states that this is calculated by finding the product of the probability of y occurring and the probability of all the feature vectors given y all divided by the probability of all the feature vectors. Then, by using the naïve assumption that all these feature vectors are independent this can be rearranged to the following line which states that the probability of a feature y given x_1 through x_n given any feature x_i is equal to the probability of feature x_i given a feature y.

Probability of y given some input sequence of events is the probability of y multiplied by the sum of each individual vector probability occurring given y occurring. This is then all divided by the probability of the whole sequence of vectors occurring. As the probability of all the vectors occurring is constant due to the nature of the input data the next lines that follow can be applied. The next line states that the probability of y occurring given a sequence of vectors (x_1 through x_n) – the overall aim of classification - is proportional to the probability of y multiplied by the sum of each individual vector given y occurring each time. Finally, the estimated value of y can be calculated by applying the *argmax* function to the right side of this equation. 'The argmax function returns the argument for the target function that returns the maximum value from the target function (Brownlee, 2020)'. This last line described here is known as the maximum likelihood function.

To estimate P(y) and $P(x_i|y)$ Maximum A Posteriori (MAP) estimation can be used. This is done using the formula shown in equation 4 for MAP (Horii, 2020):

$$\hat{\theta}_{MAP} = argmax_{\theta}\{P(\theta|D)\} = argmax_{\theta}\{\frac{P(D|\theta)P(\theta)}{P(D)}\} = argmax_{\theta}\{P(D|\theta)P(\theta)\}$$

Equation 4

MAP estimation estimates the value that maximises the posterior probability – $P(D|\theta)$. MAP is based on the same argmax formula as explained above.

For the experiments in this report the type of Naïve Bayes Classifier being used is a Gaussian Naïve Bayes Classifier. Gaussian Naïve Bayes classification works by calculating the mean and standard deviation of each input class. This model also adopts the *Gaussian Assumption* which assumes that the data supplied is of a Gaussian distribution (Shin, 2020). The likelihood of any given feature is calculated using equation 5, shown below.

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

Equation 5

The parameters σ_y and μ_y are estimated using maximum likelihood (the same technique used in Logistic Regression models).

The proof shown in figure 5 shows the theory behind Bayes Theorem which essentially states the formula shown in equation 6:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Equation 6

This theorem describes the product of the proof in figure 5. It states that the probability of a vector A occurring given be is equal to the probability of a vector B occurring given vector A multiplied by the probability of vector A, all then divided by the probability of vector B occurring. This is the basis for the Gaussian Naïve Bayes model implemented for this project (*1.9. Naive Bayes — scikit-learn 0.24.1 documentation*, 2021).

The reasons for including a Naïve Bayes Classification model are twofold; firstly, it is very fast compared to more complex models (e.g., Neural Networks) and secondly, because they do not require a huge amount of data unlike many other ML models. Furthermore, since each feature class can be independently estimated as its own dimensional distribution (and is not conditionally dependent on any other feature class) it removes the problems created by the curse of dimensionality ('Curse of dimensionality', 2021).

3.2.3 Naïve Bayes Model Data Pre-Processing

For the final Naïve Bayes classification model, normalisation and SMOTE were applied. Features with more than 2000 missing values were also removed.

4. Results

In the following section results will be analysed using a variety of metrics. Scikit-Learn's classification report uses the following measures to evaluate the performance of ML models: precision, recall and F1-score. Precision is defined as the 'fraction of relevant instances among the retrieved instances i.e., it is the total number of true positives divided by the sum of the true positives and false positives. Recall is defined as the 'fraction of relevant instances that were retrieved i.e., it is the number of true positives divided by the sum of the true positives and false negatives. The F1-score is a 'function of precision and recall' (Shung, 2020) and is calculated by the product of precision and recall by the sum of precision and recall all multiplied by 2 as shown in equation 7 below. F1-score is often favourable over precision and recall alone since it provides a balance between precision and recall. For the purposes of this research, it is also favourable over accuracy (which is simply a measure of all the correctly identified cases) since it gives a far better indication of the incorrectly classified cases and a reasonable balance between precision and recall. This is important for this research since the classes are extremely imbalanced (Huilgol, 2019) (there are far more 0 cases than 1). If only accuracy was looked at this would be misleading as there are many correclty identified 0 cases but not many correctly identified 1 cases meaning accuracy could be extremely high despite the fact that the model is not actually very useful.

$$F1 = 2 \times \frac{Precision * Recall}{Precision + Recall}$$

Equation 7

Other metrics that are worth considering are the macro and weighted averages that Scikit-Learn's classification report also produces data on. Macro-average is the mean of the precision, recall and F1-score of all the classes. Weighted average is the total number of true positives of all classes divided by the total number of datapoints across all classes (Urbanowicz and Moore, 2015).

The overall, results are be presented firstly, on their performance without oversampling techniques, and then with oversampling techniques to show how beneficial oversampling can be on unbalanced datasets.

Initial results for the overall performance across all models when predicting the occurrence of banking crises is strong (all models achieving 85%+ weighted average score). However, upon closer analysis this is misleading. Within the Adjusted MPF dataset there are 4923 datapoints where a financial crisis did not occur (value of 0) and 477 datapoints where a financial crisis did occur (value of 1). The dataset is therefore extremely imbalanced with less than 9% of the datapoints representing instances where a financial crisis has occurred. Consequently, the models have plenty of data for instances where crises do *not* occur and can therefore predict these instances with extremely high accuracy (90%+ in most cases). However, when it comes to accurately predicting instances where financial crises have occurred all models perform poorly achieving a precision score of less than 10% in most cases. Hence, the weighted average mentioned above producing a misleading accuracy score (since the weighted average score is simply the mean of the precision score for the 0 class and the precision score of the 1 class).

To address this issue of imbalanced classes oversampling was used (as discussed in the methodology section) and this led to significant improvements in the KNN and Naïve Bayes models performance on the 1 class as shown in the following subsections.

4.1 Logistic Regression Results

As shown in *figure 6,* the Logistic Regression model with normalisation has a high weighted average of 89% and, for the 0 class, a high precision score of 92%. However, it has terrible precision score of 0 for the 1 class (instances where a crisis did occur). This means it never managed to predict the occurrence of a crisis (due to the afore-mentioned reasons).

	precision	recall	f1-score	support
0	0.92	1.00	0.96	1496
1	0.00	0.00	0.00	124
accuracy			0.92	1620
macro avg	0.46	0.50	0.48	1620
weighted avg	0.85	0.92	0.89	1620

Figure 6: Classification Report of Logistic Regression Model with Normalisation

This poor performance is more accurately depicted in *figure 7* which shows the Receiver Operator Characteristic (ROC) graph for the Logistic Regression model. Good models have a large Area Under Curve (AUC) and the closer the AUC is to 0.5 the worse the model is. This model has an AUC of 0.59 which is not good.

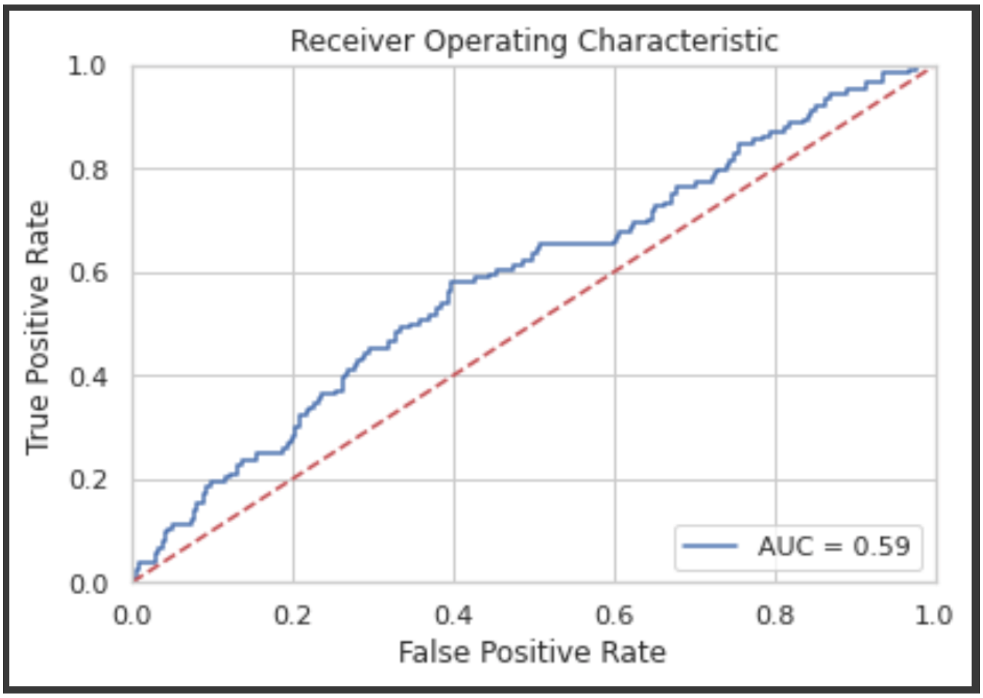

Figure 7: ROC Curve for Logistic Regression Model with Normalisation

As mentioned in the methodology section, PCA was experimented with for the Logistic Regression model, with little success. It ended up removing too many dimensions resulting in the PCA Logistic Regression model suffering from the imbalanced classes and therefore, performing very poorly in the 1 category, as shown in figure 8. It had an average f1-score of just 0.01.

	precision	recall	f1-score	support
0	0.91	0.99	0.95	2461
1	0.06	0.01	0.01	239
accuracy			0.90	2700
macro avg	0.49	0.50	0.48	2700
weighted avg	0.84	0.90	0.87	2700

Figure 8: Classification Report for Logistic Regression Model with PCA

4.2 K-Nearest Neighbors Results

As shown in *figure 9,* the KNN model performed slightly better than the Logistic Regression model across the two classes with a weighted average of 88% and a precision score of 92% in the 0 class and 38% in the 1 class. This clearly shows that, like the Logistic Regression model, the KNN model is good at predicting instances where financial crisis does not occur. However, where it differs from the Logistic Regression model is for instances where banking crises *do* occur. Here, as shown in *figure 10* it achieves a precision of 38%. Although this is still not great it is a considerable improvement of the precision of the Logistic Regression model when it comes to the 1 class.

	precision	recall	f1-score	support
0	0.92	0.97	0.94	981
1	0.33	0.16	0.22	99
accuracy			0.89	1080
macro avg	0.62	0.56	0.58	1080
weighted avg	0.87	0.89	0.88	1080

Figure 9: Classification Report for KNN Model without Normalisation

	precision	recall	f1-score	support
0	0.92	0.97	0.95	982
1	0.38	0.19	0.26	98
accuracy			0.90	1080
macro avg	0.65	0.58	0.60	1080
weighted avg	0.87	0.90	0.88	1080

Figure 10: Classification Report for KNN Model with Normalisation

Figure 11 shows the classification report for SMOTE on KNN using the borderline 1 technique. It shows an average f1-score of 0.8 which, although is lower than the average f1-score of KNN without SMOTE, when looking at the scores for the 0 and 1 classes individually it can be seen that, for the 1 class, the f1-score has increased by over 45%. However, it has also decrease by approximately 6% in the 0 class. Despite this decrease in accuracy in the 0 class the SMOTE model is a better overall model than KNN without SMOTE. This is because it has a better performance across the two classes. KNN with SMOTE is very good at predicting when there will not be a crisis. However, it's f1-score for the 0 class is so close to 1 that it suggests the model is probably overfitted to the Adjusted MPF dataset such that would not prove useful when predicting future crises based on unseen data. The fact that the SMOTE KNN model has a much higher f1-score in the 1 class combined with the more realistic f1-score of 0.88 in the 0 class suggests that it is less overfitted to the MPF dataset and would perform much better in real world unseen crises prediction instances.

smote with borderline1

	precision	recall	f1-score	support
0	0.96	0.82	0.88	995
1	0.23	0.62	0.33	85
accuracy			0.80	1080
macro avg	0.59	0.72	0.61	1080
weighted avg	0.90	0.80	0.84	1080

Figure 11: KNN with SMOTE Borderline1 Classification Report

KNN with SMOTE using the borderline 2 technique has a similar set of overall results to the borderline 1 technique (as shown in the figure 12) but is not quite as effective. It sees a larger fall in accuracy for the 0 class and a smaller increase in accuracy for the 1 class when compared with the improvements seen with regards to the borderline 1 method.

smote with borderline2

	precision	recall	f1-score	support
0	0.96	0.75	0.84	995
1	0.18	0.62	0.28	85
accuracy			0.74	1080
macro avg	0.57	0.69	0.56	1080
weighted avg	0.90	0.74	0.80	1080

Figure 12: KNN with SMOTE Borderline 1 Classification Report

The KNN model with SMOTE using the SVM technique is the best performing KNN model. As shown in figure 13 it has a higher overall accuracy of 0.83, when compared with the borderline 1 method (which had an accuracy of 0.8). It also has a very high accuracy of 0.9 in the 0 class and highest accuracy of all the models in the 0 class – achieving 0.36. This suggests it is the best performing overall model.

```
smote with svm - not overfitted - not predicitng all values of 0
correctly & prediction of 1s has also improved

              precision    recall  f1-score   support

           0       0.96      0.85      0.90       995
           1       0.26      0.60      0.36        85

    accuracy                           0.83      1080
   macro avg       0.61      0.73      0.63      1080
weighted avg       0.91      0.83      0.86      1080
```

Figure 13: KNN with SMOTE Using SVM Classification Report

Finally, the last sampling approach used on the KNN models was ADASYN. As shown in figure 14 it generally, has improved accuracy than the standard KNN models. However, it does not perform quite as well as the KNN model with SVM SMOTE applied to it.

```
ada - classificaiton report

precision    recall  f1-score   support

           0       0.97      0.78      0.86       995
           1       0.21      0.69      0.32        85

    accuracy                           0.77      1080
   macro avg       0.59      0.74      0.59      1080
weighted avg       0.91      0.77      0.82      1080
```

Figure 14: KNN with ADASYN Classification Report

4.3 Naïve Bayes Classification Results

The Naïve Bayes classifier, as shown in *figure 15,* has a high precision for the 0 class of 92%. As with the KNN model the precision score for the 1 class is considerably better than that of the Logistic Regression model, at 14%. However, it is also considerably worse than that of the Naïve Bayes Classifier.

```
              precision    recall  f1-score   support

           0       0.97      0.16      0.28       982
           1       0.10      0.95      0.18        98

    accuracy                           0.23      1080
   macro avg       0.54      0.56      0.23      1080
weighted avg       0.89      0.23      0.27      1080
```

Figure 15: Classification Report for Naïve Bayes Classifier without Normalisation

	precision	recall	f1-score	support
0	0.92	0.94	0.93	986
1	0.14	0.10	0.11	94
accuracy			0.87	1080
macro avg	0.53	0.52	0.52	1080
weighted avg	0.85	0.87	0.86	1080

Figure 16: Classification Report for Naïve Bayes Classifier with Normalisation

Naïve Bayes with plain SMOTE did not prove to be very effective (as shown in figure 17). It's f1-score decreased when compared to the model that did not use SMOTE. This model was therefore largely disregarded.

	precision	recall	f1-score	support
0	0.92	0.70	0.80	979
1	0.13	0.44	0.20	101
accuracy			0.68	1080
macro avg	0.53	0.57	0.50	1080
weighted avg	0.85	0.68	0.74	1080

Figure 17: Naive Bayes with Plain SMOTE

4.4 Overall Results

As shown from the above results it can be said that for this investigation the KNN model that uses SMOTE with SVM is the best performing model. With a high average f1-score of 0.83 and the highest f1-score the 0 class and a strong f1-score of 0.9 in the 0 class.

However, these results are slightly misleading. Firstly, because their performance when looking just at the 1 class is far lower. Secondly, the models have not had much post-processing performed on them. Consequently, they could be very overfitted for the 0 class and perhaps underfitted on the 1 class. Cross validation could have been used to discover if this is true (see discussion and future work section).

In conclusion, it is unlikely the models created here would be of any use in real-world scenarios (i.e., they will not be able to effectively predict future crises in any meaningful timeframe). However, the experiments conducted here could provide a foundation for further research into this area and the data concerning it (see discussion and future work section).

5. Discussion and Future Work

Despite seeing significant improvements to all models as more types of experiments were run (e.g., by implementing scaling, normalisation, and sampling techniques), it is still clear there is significant room for improving the models when it comes to their accuracy on the 1 class. The initial point of implementing these models was to accurately predict the occurrence of crises. Since the highest accuracy score when it came to the predicting crisis (1 class) was just 0.36 there is likely a lot more that could be experimented with to give the models that are predicting this class a much higher accuracy.

With more time a wider range of models and sampling techniques could be implemented to see if accuracy could be made higher in the 1 class. Due to the relatively low accuracy on the 1 class, it is unlikely that the models implemented in this project would be able to predict the occurrences of crises more effectively than existing ML techniques.

Another possible way of improving model performance would be to find or create a better-quality dataset that had a broader time range with more instances from the 1 class to have a more balanced dataset. However, due to the nature of the data being analysed (the occurrences of financial crises) this may prove difficult as there have only ever been a fixed number of financial crises. The nature of the datasets required for financial crisis prediction raises the fundamental question of whether ML models are the best suited ways of predicting the occurrence of a crisis. Since the most effective ML models generally require very large datasets to be trained effectively, it could be said that econometrics methods are better suited for this type of classification task. The reason many ML models (e.g., Convolutional Neural Networks) were not even considered for this classification task was because they require datasets that have far more information to be trained on than were available for this project. For example, CNNs tend to be very well suited for image classification were there are huge amounts of data stored in each individual image and many images can be fed into the models.

To establish whether it was model design or the fundamental quality of data that led to poor accuracy in the 1 class further work could be done. This could involve finding better sources of more complete data. It could also involve designing better models, perhaps with more pre- and post-processing techniques on the data. Pre-processing that could benefit the dataset could include under sampling the 0 class whilst also oversampling the 1 class to further balance the data. Regularisation and dropout techniques could also be used to further avoid the issue of the model being overfitted. With regards the fundamental model used, a random forest classifier could be experimented as well.

Additional techniques that could have been experimented with more time and resources could have been post-processing techniques such as a k-Fold Cross Validation to check if the models were overfitted.

With an abundance of time, a website would have also been made to show the results online and make them even easier for the public to access.

Other areas of research that could be investigated with more time and resources could be related to the correlation of ESG indicators with traditional economic instruments such as sovereign CDS information. This was a secondary objective of this report and was not done due to a lack of sufficient time. Experiments that could have been done would have been firstly on analysis of ESG features that had the most influence when predicting economic events. These specific features could have then been used to see if there would be a better way to price CDS than the current ways. If done successfully this could provide an invaluable model for global economic markets and governmental departments.

6. Conclusion

In conclusion it is evidently difficult to accurately predict the occurrence of financial crisis events – it has been tried using a huge range of methods to varying degrees of success in the past – and anyone that could predict future crises with a high degree of accuracy would have found a revolutionary model. This extreme difficulty of the central objective of this project then raises the issue; was this the right central objective? It is likely that more useful results would be obtained in the focus of the experiments had not been on a binary classification task but on trying to focus on features that may increase the risk of certain economic events by certain proportions. In the end economic uncertainty is vast and will always be. Therefore, in hindsight, it would make more sense to focus on particular areas of ESG indicators that may contribute to higher chances of significant financial events occurring.

Overall, despite the final models not being that effective at predicting when financial crises would occur there were various techniques – particularly normalisation and SMOTE – which helped improve accuracy by significant proportions.

Although oversampling did help to address the imbalanced dataset that was being used it was only able to help to a certain extent. Ultimately, although improvements to the model were achieved, it was likely not enough to make the model be of significant use in real-world crisis prediction situations. This is likely down to either the fundamental design of the models which could be improved with more time and resources or a different approach.

An alternative approach that could improve the overall model would be to apply more effective feature selection on the ESG databased combined with traditional measures of financial performance such financial times series to create a more well-rounded dataset.

At the time of writing there is very little public research into the effect of ESG indicators on large scale financial crisis concerned with nations. There is, however, a significant amount of research into the correlation between company's investment and commitment to strong ESG performance and their own financial success (Caterina De Lucia, Pasquale Paziena, and Mark Bartlett, 2020) (Broadstock *et al.*, 2021).

Despite the models created in this research project not being effective for future, real-world scenarios, the research could provide the groundwork for future analysis of this domain and the Adjusted MPF data's relationship with traditional economic events. The task of figuring out when the next crisis will occur is one of much debate in the global financial community and attempting to do it with modelling techniques such as the ones used in this investigation alone is likely extremely difficult or impossible.

7. Bibliography

1) *1.6. Nearest Neighbors — scikit-learn 0.24.1 documentation* (2021). Available at: https://scikit-learn.org/stable/modules/neighbors.html#classification (Accessed: 26 April 2021).

2) *1.9. Naive Bayes — scikit-learn 0.24.1 documentation* (2021). Available at: https://scikit-learn.org/stable/modules/naive_bayes.html#gaussian-naive-bayes (Accessed: 26 April 2021).

3) *6.3. Preprocessing data — scikit-learn 0.24.2 documentation* (2020). Available at: https://scikit-learn.org/stable/modules/preprocessing.html#preprocessing-scaler (Accessed: 11 May 2021).

4) *ADASYN — imbalanced-learn 0.3.0.dev0 documentation* (2020). Available at: http://glemaitre.github.io/imbalanced-learn/auto_examples/over-sampling/plot_adasyn.html (Accessed: 3 May 2021).

5) Antoniades, A., Widiarto, I. and Antonarakis, A. S. (2019) 'Financial crises and the attainment of the SDGs: an adjusted multidimensional poverty approach', *Sustainability Science*. doi: 10.1007/s11625-019-00771-z.

6) *BCS Code of Conduct | BCS* (2020). Available at: https://www.bcs.org/membership/become-a-member/bcs-code-of-conduct/ (Accessed: 18 May 2021).

7) Bhattacharyya, I. (2018) *SMOTE and ADASYN (Handling Imbalanced Data Set), Medium*. Available at: https://medium.com/coinmonks/smote-and-adasyn-handling-imbalanced-data-set-34f5223e167 (Accessed: 11 May 2021).

8) Bluwstein, K. *et al.* (2020) 'Credit Growth, the Yield Curve and Financial Crisis Prediction: Evidence from a Machine Learning Approach', *SSRN Electronic Journal*. doi: 10.2139/ssrn.3520659.

9) Broadstock, D. C. *et al.* (2021) 'The role of ESG performance during times of financial crisis: Evidence from COVID-19 in China', *Finance Research Letters*, 38, p. 101716. doi: 10.1016/j.frl.2020.101716.

10) Brownlee, J. (2020) 'What Is Argmax in Machine Learning?', *Machine Learning Mastery*, 2 April. Available at: https://machinelearningmastery.com/argmax-in-machine-learning/ (Accessed: 10 May 2021).

11) Caprio, G. and Klingebiel, D. (1999) *Bank Insolvencies: Cross-Country Experience*. The World Bank (Policy Research Working Papers). doi: 10.1596/1813-9450-1620.

12) Caterina De Lucia, Pasquale Pazienza, and Mark Bartlett (2020) 'Does Good ESG Lead to Better Financial Performances by Firms? Machine Learning and Logistic Regression Models of Public Enterprises in Europe'.

13) Chawla, N. V. *et al.* (2002) 'SMOTE: Synthetic Minority Over-sampling Technique', *Journal of Artificial Intelligence Research*, 16, pp. 321–357. doi: 10.1613/jair.953.

14) Chen, J. (2020) *Yield Curve*, *Investopedia*. Available at: https://www.investopedia.com/terms/y/yieldcurve.asp (Accessed: 9 May 2021).

15) 'Commonly Used Machine Learning Algorithms | Data Science' (2017) *Analytics Vidhya*, 8 September. Available at: https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/ (Accessed: 18 May 2021).

16) 'Credit default swap' (2021) *Wikipedia*. Available at: https://en.wikipedia.org/w/index.php?title=Credit_default_swap&oldid=1018048208 (Accessed: 9 May 2021).

17) Crifo, P., Diaye, M.-A. and Oueghlissi, R. (2015) 'Measuring the effect of government ESG performance on sovereign borrowing cost'. Available at: https://hal.archives-ouvertes.fr/hal-00951304 (Accessed: 11 May 2021).

18) 'Curse of dimensionality' (2021) *Wikipedia*. Available at: https://en.wikipedia.org/w/index.php?title=Curse_of_dimensionality&oldid=1009164221 (Accessed: 26 April 2021).

19) Dorpe, S. V. (2018) *Preprocessing with sklearn: a complete and comprehensive guide*, *Medium*. Available at: https://towardsdatascience.com/preprocessing-with-sklearn-a-complete-and-comprehensive-guide-670cb98fcfb9 (Accessed: 11 May 2021).

20) 'Euclidean distance' (2021) *Wikipedia*. Available at: https://en.wikipedia.org/w/index.php?title=Euclidean_distance&oldid=1017589549 (Accessed: 10 May 2021).

21) He, H. *et al.* (2008) 'ADASYN: Adaptive synthetic sampling approach for imbalanced learning', in *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence). 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 1322–1328. doi: 10.1109/IJCNN.2008.4633969.

22) He, H. and Ma, Y. (eds) (2013) *Imbalanced Learning: Foundations, Algorithms, and Applications*. 1st edition. Hoboken, New Jersey: Wiley-IEEE Press.

23) Horii, S. (2020) *MLE, MAP and Bayesian Inference*, *Medium*. Available at: https://towardsdatascience.com/mle-map-and-bayesian-inference-3407b2d6d4d9 (Accessed: 10 May 2021).

24) Huilgol, P. (2019) *Accuracy vs. F1-Score*, *Medium*. Available at: https://medium.com/analytics-vidhya/accuracy-vs-f1-score-6258237beca2 (Accessed: 10 May 2021).

25) *Index Suite* (2020). Available at: https://www.jpmorgan.com/insights/research/indices/product (Accessed: 18 May 2021).

26) Kalamara, E. *et al.* (2020) 'Making Text Count: Economic Forecasting Using Newspaper Text', *SSRN Electronic Journal*. doi: 10.2139/ssrn.3610770.

27) '*k*-nearest neighbors algorithm' (2021) *Wikipedia*. Available at: https://en.wikipedia.org/w/index.php?title=K-nearest_neighbors_algorithm&oldid=1018865836 (Accessed: 3 May 2021).

28) Kouiroukidis, N. and Evangelidis, G. (2011) 'The Effects of Dimensionality Curse in High Dimensional kNN Search', in *2011 15th Panhellenic Conference on Informatics. 2011 15th Panhellenic Conference on Informatics*, pp. 41–45. doi: 10.1109/PCI.2011.45.

29) Laeven, L. and Valencia, F. (2018) 'Systemic Banking Crises Revisited', *IMF Working Papers*, 18(206), p. 1. doi: 10.5089/9781484376379.001.

30) 'Logistic function' (2021) *Wikipedia*. Available at: https://en.wikipedia.org/w/index.php?title=Logistic_function&oldid=1020992269 (Accessed: 10 May 2021).

31) Margaretic, P. and Pouget, S. (2018) 'Sovereign bond spreads and extra-financial performance: An empirical analysis of emerging markets', *International Review of Economics & Finance*, 58, pp. 340–355. doi: 10.1016/j.iref.2018.04.005.

32) Pacca, L. *et al.* (2020) 'The effect of financial crises on air pollutant emissions: An assessment of the short vs. medium-term effects', *Science of The Total Environment*, 698, p. 133614. doi: 10.1016/j.scitotenv.2019.133614.

33) Rockström, J. *et al.* (2009) 'A safe operating space for humanity', *Nature*, 461(7263), pp. 472–475. doi: 10.1038/461472a.

34) *SDG Indicators — SDG Indicators* (2017). Available at: https://unstats.un.org/sdgs/indicators/indicators-list/ (Accessed: 7 May 2021).

35) Shin, T. (2020) *A Mathematical Explanation of Naive Bayes in 5 Minutes*, *Medium*. Available at: https://towardsdatascience.com/a-mathematical-explanation-of-naive-bayes-in-5-minutes-44adebcdb5f8 (Accessed: 10 May 2021).

36) Shung, K. P. (2020) *Accuracy, Precision, Recall or F1?*, *Medium*. Available at: https://towardsdatascience.com/accuracy-precision-recall-or-f1-331fb37c5cb9 (Accessed: 10 May 2021).

37) *sklearn.linear_model.LogisticRegression — scikit-learn 0.24.1 documentation* (no date). Available at: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html (Accessed: 25 April 2021).

38) *sklearn.preprocessing.normalize — scikit-learn 0.24.1 documentation* (2020). Available at: https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.normalize.html (Accessed: 25 April 2021).

39) Urbanowicz, R. J. and Moore, J. H. (2015) 'Metrics and Scoring: quantifying the quality of predictions', *Evolutionary Intelligence*, 8(2–3), pp. 89–116. doi: 10.1007/s12065-015-0128-8.

40) *Welcome | Environmental Performance Index* (2021). Available at: https://epi.yale.edu/ (Accessed: 25 April 2021).

41) World Bank (2020) *Worldwide Governance Indicators*. Available at: https://databank.worldbank.org/source/worldwide-governance-indicators.

42) *World Development Indicators | DataBank* (2021). Available at: https://databank.worldbank.org/source/world-development-indicators (Accessed: 25 April 2021).

8. Appendices

8.1 Project Logs

8.1.1 Meeting 1st February 2021

If country is doing particularly well – can relate that too financial variables – people want a product that can do this

Read:

- Black rock – Stanley Fischer – ESG annual report

Rachel work:

- Look at typical approach – many ways to 'slice the pie'
- SDG – social development goals
- Look at things that show how well country is developing e.g., poverty, water availability etc.
- Rachel's work is more focused on macro social indicators e.g., inequality – wealth, income, gender etc. – lots of papers relating this to success of a country/company
- Board's with more diversity – tend to do much better – if u can prove this you can start to make statements – becomes political
    - Rachel relates back to the CDS data – is that country priced like a risky country (high yields)
    - E.g., bad on poverty/water – have been doing well recently – maybe that
- Data is publicly available – 3/4 companies (private companies claiming they can provide better quality data)
- Data sources:
    - **World Bank data portal on ESGs**
        - Good aggregator of data – **have an API you can write to**
    - **United Nations**
        - Phrases: SDGs
    - Laeven & Valencia paper data stops post financial crisis – will be a problem
    - CDS data is from 2005 latest
    - Nature of this data is very slow – *don't need to worry about last two years*
    - 2 main components –
        1. Can you predict financial crises with ESG data?
        2. Are the CDSs priced correclty based on ESG indicators

Think about:

- CDS rates – the riskiness of investing in a company – tradeable instruments
- CDS is clearly going to be influenced by ESG things
- Another factor:
    - Political events e.g., coup in Myanmar – **look into this** – did ESG factors indicate this?

    - CDS price will be massively influenced by a country's economy – include a few high level general economic indicators
        - Will find some things
        - Can think of governance as the wrapper that holds everything together
    - Find a paper by a credit rating agency – see what they say about governance – in what ways is governance an important factor – biggest credit rating agencies:
        - *Standard & Poor's*
        - *Moody's*
        - *Fitch*
- Will get a different take if you go to
- Look for MSCI index
    - https://www.msci.com/index-solutions
    - https://www.msci.com/esg-investing
    - Global index of markets across the world – they create an ESG product that they sell
    - Gives a report into your investment strategy to evaluate how well they perform against ESG indicators – and gives them alternative ideas that are more *ethical*
- Approach to build the model – If you want advanced ML techniques:
    - *Panel regression – start with this* – amount of data I actually have makes complex models more challenging to work with
        - *Start with regressions & basic correlation plots*
    - Rachel is amazed – most people don't do exploratory analysis – lots of people build complex models with data that is not appropriate for the models it is being fed in to
    - Some data will be binary (high/low)
    - GDP will be a scale
    - Need to *perform intelligent feature selection* → once this has been done it *may be a simple regression model* – this is complex enough (in terms of complexity required for final year projects) if you have done good pre-processing and visualisations
    - From practitioner's perspective – reason to believe countries behave similarly due to various different factors e.g., religion, climate, politics
        - Another possible component – look into *similar behaviour according to regions* → Could try a clustering approach
        - E.g., couple of principle components, k-clustering etc. removing outliers
            - E.g., these countries have similarities according to feature space and then make a model for those regions
- Time series element of things might be new to me – once up and running with data can ask Rachel for more help on this
- If interested in financial markets – read this:
    - Reply to email Rachel sent me with the papers – she will send more once I have done this – done
    - Look at CountryRisk.io - risk analysis using ESG data – find their paper and maybe use the site – they have the data – they do some things badly though – insight to how people do stuff and maybe help me find direction I want to take with my project

    - Pipeline analysis by March time

8.1.2 Meeting 2nd April 2021

Strength of correlation over different times for cds

Look at news archives to find out what year's crises hit

Cds 1 year prior to crisis

Feed CDS into logistic regression model & see what confusion matrix/roc curve would look like

- Take different threshold of cds & make prediction based on the CDS

Go back over leaven & Valencia paper as well

Principle component analysis of ANOTIADES ET AL. poverty framework for pre-processing

- Scale all features
- Make decision
- LOOK AT FCES MULTI POVERTY SPREADSHEET- HAS DATA ON OCCURRENCE OF CRISES
- Plot histograms/scatter plots of each of different features
- Look back over
- Follow papers – do converse of what these guys did
- Get some PLOTS!
- Endless tweaking with research projects!!

8.1.3 Meeting 16th April 2021

Do literature review & discuss methods of other papers that did similar stuff – find Rachels papers

Intro
Aims & objectives – 1st sentence/first paragraph – towards end of introduction
    Might just have list of hypotheses – these are the experiments finding these
Lit review
Methods
Results
Conclusion – have you achieved objectives – if not why not – alternative approaches that could have worked better in hindsight – with more time I would have done an extension – have still insight into topic area
    Not a PHd!

CDS data:

- Could ESG info help you to price CDS better
    - Logistic regression of CDS
    - Obtain instances of CDS in year 0 and financial crisis in year 1

    - Could do basic logistic regression of CDS to predict the FC
    - Use Rachel's formula to calculate probability of default from CDS
    - Add in ESG to see if it proves
    - New pricing strategy for CDS – could you come up with a better CDS pricing strategy using ESG data?
  - There is a formula to calculate the probability of default based on the CDS price – check that formula – email from Rachel