

Implementing NLP techniques and data for Europe financial market forecasting

Yihong Su^{1,5,9,†} Yuanhong Xiang^{2,6,†} Yiming Lai^{3,7,†} Rui Liu^{4,8,†}

¹ Haide College, Ocean University of China (Laoshan Campus), Qingdao, Shandong Province, 266100, China

² Department of Business, University of New South Wales, Sydney, NSW 2052, Australia

³ Chongqing No.1 Secondary School, Chongqing, China

⁴ University of Washington, Seattle, WA, United States

⁵ syh1119@stu.ouc.edu.cn

⁶ xiangyuanhong8@gmail.com

⁷ online051125@gmail.com

⁸ rl67@uw.edu

⁹ corresponding author

[†] These authors contributed equally

Abstract. The large military conflict between Russia and Ukraine has consequences for the global financial market. The European region is one of the hardest hits, which experienced a high energy cost and devaluation of the currency. This paper investigates the Employment Expectations Indicator (EEI) and Economic Sentiment Indicator (ESI) from the business consumer survey index to reflect the reaction of the consumers of the European market by implementing Natural Language Processing (NLP) techniques and machine learning. In the object, the monetary policy decisions from European Central Bank are considered the data set for forecasting the Index. In the result, all nine models from 'scikit-learn' contributed a great job in both categories of Accuracy, Precision, and F-score and Recall with a level of 90% as average. It represents that the NLP techniques effectively forecast the future value of two European business and consumer indexes. However, the limit of the dataset and model mostly related to the performance of NLP techniques is working for forecasting but less related to the expected values of the EEI and ESI indices. Future research would focus more on forecasting the expected value of the two indexes and reflect the consumers' reactions in the markets.

Keywords: NLP techniques, machine learning, deep learning, financial market prediction.

1. Introduction

In February of 2022, a large-scale military conflict between Russia and Ukraine Kicked off the prologue, which shocked the rest of the world's minds with its intensity. The countries led by the United States and the European Union quickly implemented financial sanctions to cripple the Russian economy, including seizing Russian property held overseas. In addition, energy exports were limited, and the ban of certain imports into Russia. The West's advanced aversion to Russia's military activity directly exiled

a country that is one of the biggest exporters and providers in the global energy market. It directly resulted in severe consequences for the global financial market.

The absence of Russia has directly disrupted the global market in several significant aspects. For starters, without Russia, the energy market's supply chain, with its discontinuity and high demand, directly breaks the EU-Russia market interdependence and oil dependency relationship, which led to an increase in the price of crude oil [1]. The price of crude oil surged from \$65.44 on December 1, 2021, to \$111.60 on March 22, 2022, and peaked on March 8, 2022, at \$123.7 per barrel with an upsurge of 86.67% [2]. The extremely high price of crude oil directly impacts global inflation. Inflation is an increase in the price level of household goods and services, measured by the rate of change of these prices [3]. The inflation level of major European countries published by the European Central Bank (ECB), the inflation level for the Euro area rose to 10.6% on October 20, 2022, the highest level since 1997. Otherwise, the rising inflation level also represents the weakening of the euro's purchasing power.

However, because energy commodity markets are global, changes in the price of oil in one part of the world ultimately affect global oil prices, including in the United States, which has a low level of dependence on commodities from Russia [4]. The inflation index in June 2022 achieved 9.06%, the highest level in the last 40 years [5]. Furthermore, it is a sign that U.S. inflation is out of control, which adds to the country's uncertainty. To combat the uncertainty, the Fed (The Federal Reserve, the United States central bank) implemented a new round of monetary contraction policy to keep fighting inflation. However, tighter U.S. monetary policy squeezes economic activity almost everywhere by curbing risk appetite and pushing up the dollar's value [6]. The countries of the European region are one of the hardest hits; they not only have to accept the shortage and high cost of energy resources caused by the absence of Russia but also must be against the devaluation of their currencies because of the U.S.'s contractionary policy. A new round of global recession is also looming.

The objective of this paper is to utilize Natural Language Processing (NLP) techniques and machine learning algorithms to examine the impact of the contractionary policy from the U.S. and the Russo-Ukrainian War on the European financial market. To achieve this goal, we incorporate "Business and Consumer Surveys" indicators as means to analyze the sentiment and response of the public to this exogenous event [7]. We also assess the accuracy of the NLP techniques in predicting the trends of these indices, which reflect the genuine reactions of consumers in the financial markets. Our approach provides a comprehensive analysis of the impact of the external events on the European financial market by considering the social and behavioral factors that influence market dynamics. The findings of this study offer valuable insights for policymakers and stakeholders in the financial sector to make informed decisions based on the reactions of the market participants.

The Business and consumer surveys indicators - The Directorate General for Economic and Financial Affairs regularly conducts harmonized surveys for several economic sectors in the European Union and in the candidate countries, which include – *Industrial Confidence Indicator* (INDU), *Services* (SERV), *Consumer* (CONS), *Retail Trade* (RETA), *Construction* (BUIL), *Economic Sentiment Indicator* (ESI) and *The Employment expectations indicator* (EEI) [7]. The last two indicators are the focus point in this paper.

2. Literature review

2.1. The methods and results of previous research

There is much comparative research and previous studies about financial market forecasting and the relationship between the factors and fluctuation of the financial markets.

Researchers have used daily panel data from 12 chosen nations that spanned four continents to assess the pandemic's sharply adverse effects on stock market performance [8]. The research then employs the event study methodology while measuring the effect on economic activity using a panel vector autoregressive model. They discovered that in comparison with other markets, European markets suffered the most. In addition, all stock markets are negatively impacted by all pandemic factors, and economic activity is negatively impacted by lockdown days and movement restrictions. Also, this study

considers those nations that substantially contribute to global economic growth and suffer severely from the Coronavirus Disease 2019 (COVID-19) epidemic. The outcome-based recommendations will help governments, regulatory authorities, and policymakers combat the crisis in different dimensions.

It was examined that the influence of the Russo-Ukrainian war on the condition of dependence on Russian commodities [9]. The research applies panel data methodology to analyze data from a large panel of 73 countries to conclude that War-related shocks have a considerable impact on the financial markets. However, asset values are less affected than volatility. Markets see reliance on Russian commodities as a significant risk element, reducing equity returns and increasing instability. There is a threshold for adverse effects on asset prices when the effect of war on returns surpasses the [0-20%] level for countries' reliance. No matter how dependent society is, armed conflict makes it more unstable; the effect grows as dependency increases. The findings have significance for international communication diversification efforts.

In the previous research, many different methodologies and approaches were used to forecast and examine information connected to the financial markets, all of them being manually intensive tasks of polling individuals. While surveys are adequate, given today's computing power, we can get into the minds of millions to capture similar dimensions provided by a manual survey process. One such field to extract opinions, sentiments, and emotions is NLP, an entire field leveraging algorithms to understand human language, and the results have been quite fascinating. Leveraging sentiment extracted from social media, specifically Twitter, successfully predicted the 2016 US election results with an 81% model accuracy [10]. Similarly, Bollen achieved an accuracy of 87.6% in predicting the Dow Jones Index using various emotional dimensions extracted from tweets, reinforcing that prediction of the stock market is possible [11].

In recent years, with the progress of technology, deep learning and sentiment analysis have become more widely used. Sentiment analysis, also known as opinion mining, aims to analyze the emotional polarity (positive, negative, neutral) of human beings' opinions towards goods, services, organizations, and events. Sentiment analysis in different scenarios may represent different technical applications such as emotion recognition, emotion classification, opinion mining, opinion analysis, opinion extraction, subjective analysis, emotion calculation, and evaluation analysis [12]. In summary, sentiment analysis analyzes human opinions about a target object. If one influencer, for example, suggested that the iPhone 15 would go on sale much later than expected. The emotion expressed in such expectation can be analyzed by NLP technique.

Time is also significant in sentiment analysis. Opinion time is also conducive to subsequent analysis, and people in different periods significantly differ in their evaluation of certain things. Hence, we should make good decisions about the duration of the data.

Social Media is a medium where people can express their thoughts freely. Social platforms are also a platform to help young people broaden their horizons. Twitter is the world's largest social media company, with 130 million Twitter accounts. Twitter has 33 billion monthly active users (MAU) [13]. These data show that people depend very much on Twitter, and the freshness of information is also excellent.

National policies shape the company's operating environment and will affect the company in various aspects: for example, national taxes, subsidies, environmental protection policies, and corporate competition policies., all of which will affect the stock development of an industry and a company. The constant change of national policies will affect the national and even the global financial environment. If a policy is predictable and in line with public perception, it will have less impact on the financial markets [14]. However, the stock market will be more volatile if it is a misguided policy. Therefore, international policy is critical, and the impact on stocks is also significant.

Since the Saunder study in 1993, more people have looked at the effect of weather on investor mood and, therefore, the economic environment [15]. Weather effects have been found in places like the United States and Thailand. Weather can also have a significant impact on specific industries. For example, a damp climate will significantly impact food and even some high-tech products. Excessive

humidity in the air may lead to a decline in instrument accuracy. Constant heat can also accelerate the corrosion of some foods. So the weather also has a significant impact on the economic environment.

In the 1950s, Alan Turing developed his famous "Turing test", which is thought to be the beginning of the idea of NLP [16]. By the 1970s, it was thought that NLP was similar to how a child learns a new language. At this time, the NLP stayed in the stage of rationalism. However, this requires researchers to be proficient in languages and computers, which is problematic. Therefore, although some simple problems have been solved, many complex problems still exist.

The next stage was from the 1970s to the early 21st century, when corpora became more and more abundant with the rapid development of the Internet, and computing power was constantly improved. The trend of thought of NLP changed from empiricism to rationalism. Jelinek and the IBM Watson Lab have been instrumental in this shift, using statistical methods to improve speech recognition rates from 70% to 90%, taking NLP from the lab to the field [17].

The third stage was from 2008 to 2019. Inspired by the achievements in image and speech recognition, people also began applying "deep learning" to NLP. They succeeded in machine translation, question-answering systems, reading comprehension, and other fields [18]. Deep learning is a multi-layer network of signals, starting from the input layer through layer by layer of nonlinear changes to obtain the output. With the input-to-output pairs ready, a neural network can be designed and trained to perform the desired task. Recurrent Neural Network (RNN) has been one of the most commonly used methods of natural language processing; Gated Recurrent Units (GRU), Long Short-Term Memory (LSTM), and other models have triggered a wave of enthusiasm.

Pre-training models have made significant progress in natural language processing in recent years. Pre-training model refers to unsupervised or self-supervised pre-training on the large-scale unsupervised corpus to acquire general language modeling and presentation capabilities [19]. When applied to actual tasks, there is no need to make significant changes to the model. Only an output layer for a specific task can be added to the original language representation model, and a little training can be performed on the model using task corpora. This step is called fine-tuning, one of the significant developments in NLP in recent years [20].

NLP is essential to us. From the words "artificial intelligence", artificial means ourselves, and "intelligence" means the machine's understanding of human language and thoughts. If machines cannot even understand human language, how can they communicate with human beings? How can they show intelligence? Therefore, NLP is a required field.

Another important field is Machine Learning, the current mainstream approach to solving the problem of artificial intelligence. It was born over one hundred years ago in the early 20th century.

2.2. Machine learning

Machine learning algorithms can be divided into supervised, unsupervised, semi-supervised, and reinforcement learning. Supervised learning learns from training samples (labeled data) to get a model and then deduces from that model. In unsupervised learning, there is no training process (unlabeled data). Given some sample data, the machine learning algorithm can directly analyze the data and obtain some knowledge. *Reinforcement learning* is a unique machine learning algorithm that determines an action based on the current state of the environment and then moves on to the next state to maximize the payoff. Semi-supervised learning can be regarded as combining supervised and unsupervised learning. Here are some important developments in supervised learning. Fisher invented linear discriminant analysis in 1936[21]. Bayesian classifiers started in the 1950s. He invented them based on Bayesian decision theory [22]. Since 1980, machine learning has become an independent director, and a hundred schools of thought have contended with rapid development [23]. In 1989, LeCun designed the first convolutional neural network for handwritten digit recognition, which is the ancestor of deep convolutional neural networks widely used today [24]. LSTM, now the most popular, appeared in 2000, integrated with a deep cyclic neural network after 2013, and achieved success in speech recognition [25]. Random Forest appeared in 2001 and has been used on a large scale in many problems [26].

Considering past researchers' success using NLP techniques, we will mainly use sentiment analysis as the primary method of data analysis. Therefore, we choose to use the Twitter database for sentiment analysis.

2.3. Deep learning

Deep learning has developed rapidly in recent years, and AlphaGo, an artificial intelligence program for Go, has also brought a wave of deep learning. In 1943, neurologist W.S. McCulloch and mathematician W. Pritts first established a neural network and mathematical model following biological neurons' structure and working principle [27]. They constructed a simplified model, which is the beginning of an artificial neural network.

In 1958, Rosenblatt proposed a neural network composed of two layers of neurons called "Perceptrons" [28]. In 1962, the method was proved to be convergent, and the theoretical and practical effects caused the first wave of neural networks [29].

However, Marvin Minsky proved in 1969 that Perceptrons could only deal with linear problems, which led to a 20-year standstill in neural network research.

In 1974, Paul Werbos of Harvard University invented the Backpropagation algorithm, but it was at a low point then, so it was mostly ignored [30]. It was not until 1986 that Geoffrey Hinton, the father of neural networks, carried forward the Backpropagation algorithm and adopted Sigmoid for nonlinear mapping, which caused the second wave of neural networks.

In 1991, the BP algorithm was pointed out to have the problem of disappearing gradient; that is to say, in the transmission of the error gradient after term, the gradient of the back layer is superimposed to the front layer in a multiplicative way [31]. Due to the saturation characteristic of the Sigmoid function, the gradient of the back layer is slight. The error gradient is almost zero when transmitted to the front layer, so learning the front layer effectively is impossible. This problem directly hinders the further development of deep learning.

In addition, in the mid-1990s, Support Vector Machine algorithms (SVM algorithm) and other shallow machine learning models were proposed [32]. SVM is also a supervised learning model for pattern recognition, classification, and regression analysis. Support vector machine is based on statistics, which differs from the neural network. The proposed algorithms, such as support vector machines, once again hinder the development of deep learning.

In 2012, Convolutional Neural Network (CNN) AlexNet emerged with several advantages. The Rectified Linear Unit (ReLU) activation function is used to increase the convergence rate and solve the problem of disappearing gradients, and the methods of pre-training and fine-tuning are abandoned. The second point is to extend the LeNet5 structure by adding the Dropout layer to reduce overfitting and the Local Response Normalization (LRN) layer to enhance generalization capability/reduce overfitting. Finally, this is the first time a Graphics Processing Unit (GPU)-accelerated model has been used.

3. Data

3.1. Monetary policy decisions and EEI & ESI indices

Both data are obtained from European Central Bank (ECB)'s official website [33-34]. The ECB is the central bank of the Eurozone, which consists of 19 European Union countries that have adopted the euro as their currency. The ECB was established in 1998, and its primary mandate is to maintain price stability in the Eurozone by setting and implementing monetary policy. To achieve this mandate, the ECB conducts regular meetings of the Governing Council, which is responsible for setting the ECB's key interest rates and making other important policy decisions. The ECB also oversees the operation of the Eurozone's payment and settlement systems and acts as a lender of last resort to Eurozone banks. Additionally, the ECB is responsible for supervising the largest banks in the Eurozone, in cooperation with national authorities. Overall, the ECB plays a critical role in ensuring the stability and soundness of the Eurozone's monetary and financial systems.

The monetary decision texts from the ECB are a valuable resource for understanding the bank's policy decisions and strategies. These texts are released after each ECB Governing Council meeting, and they provide detailed information on the bank's assessment of the economic and financial conditions in the Eurozone, as well as its policy decisions and reasoning. The texts are available in multiple languages and cover a wide range of topics, including inflation, interest rates, monetary policy operations, and financial stability. Researchers and policymakers can use these texts to gain insights into the ECB's thinking and to analyze the effectiveness of its policies. Overall, the ECB's monetary decision texts are an important tool for anyone interested in understanding the workings of the Eurozone's monetary policy.

The ECB publishes two key economic sentiment indicators, the Economic Sentiment Indicator (ESI) and the Economic Expectations Index (EEI). The ESI is a composite indicator that measures the overall economic sentiment in the Eurozone, based on surveys of businesses and consumers. The survey covers a wide range of sectors, including manufacturing, construction, services, retail, and consumer confidence. The ESI is calculated as a weighted average of the sector-specific confidence indicators, and it provides valuable insights into the overall economic outlook in the Eurozone.

The EEI, on the other hand, is a forward-looking indicator that measures the expected economic conditions in the Eurozone over the next 12 months. The EEI is also based on surveys of businesses and consumers, and it covers the same sectors as the ESI. The EEI is calculated as the weighted average of the respondents' expectations for economic activity, employment, prices, and other key economic indicators.

Both ESI and EEI are important tools for monitoring economic conditions and forecasting economic growth in the Eurozone. They are closely watched by policymakers, analysts, and investors, as they provide valuable insights into the sentiment and expectations of businesses and consumers, and can influence economic decision-making at the national and regional levels.

In total, 293 textures of ECB's monetary policy decisions ranging from March 4th, 1999 to February 2nd, 2023 are used. As for EEI/ESI, ECB provides data from Jan 31st, 1985 to Feb 28th, 2023, but only the ones that correspond to the monetary policy decisions are used.

3.2. Data pre-processing

The original texts of each ECB monetary policy were converted to the lowercase format. Next, the open-source Python NLP toolkit called Natural Language Toolkit, "nltk", was utilized to tokenize the text and remove non-letter characters. To eliminate words that do not convey much information, all stopwords were removed using `nltk.corpus.stopwords.words("english")`. The remaining words were then stemmed using nltk's SnowballStemmer to group together words that have the same meaning but are different grammatical variations of the same word. The TF-IDF score was then calculated for each token in each text, and the results were concatenated to form a matrix for further analysis.

TF-IDF, which stands for Term Frequency-Inverse Document Frequency, is a widely used technique in NLP for measuring the importance of words or terms in a document corpus. It is a statistical method that evaluates how relevant a word is to a document in a collection of documents [35].

The first part of the TF-IDF formula, Term Frequency (TF), measures the frequency of a term in a document and gives higher weight to words that appear more frequently in the document. The second part, Inverse Document Frequency (IDF), measures how common or rare a word is in the entire collection of documents, and gives higher weight to words that are rare across all documents. The IDF is calculated as the logarithm of the total number of documents in the corpus divided by the number of documents containing the term.

By combining these two metrics, TF-IDF can identify words that are both frequent within a document and unique to that document, giving them a higher importance score. This makes TF-IDF a useful tool for various NLP tasks, such as information retrieval, text classification, and document clustering.

In summary, TF-IDF is a powerful and flexible technique that can help improve the accuracy and efficiency of many NLP applications, by enabling the identification of important words and terms in large collections of text data.

Overall, these preprocessing steps aim to improve the quality and efficiency of text analysis by reducing the noise and dimensionality of the data. By converting the text to lowercase, tokenizing it, removing stopwords, and stemming the words, the resulting data set is more informative and better suited for analysis purpose. This data is shown in Table 1.

Table 1. The TF-IDF matrix of 291 texts and the tokens (sample).

Text	background	condit	monetari	conduct	latest	inform	decid	govern	meet	today
0	0.00000	0.08112	0.08580	0.14392	0.44615	0.37179	0.01923	0.03077	0.01599	0.01544
1	0.00000	0.00000	0.04892	0.00000	0.00000	0.00000	0.00000	0.03509	0.03647	0.03521
2	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.02703	0.02809	0.02712
3	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.04167	0.03333	0.03465	0.06690
4	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.03676	0.02941	0.01529	0.01476
5	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.04167	0.03333	0.03465	0.06690
6	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.06579	0.05263	0.05471	0.05281
7	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.06579	0.05263	0.05471	0.05281
8	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.04167	0.03333	0.03465	0.06690
9	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.06579	0.05263	0.05471	0.05281
10	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.06579	0.05263	0.05471	0.05281

After the text pre-processing is done, inspired by Object Oriented Programming, the analysis object was created for each text. Each object contains the preprocessed text, its TF-IDF vector, and the matching EEI and ESI (these two indices are simply called indices below). As said above, the indices are released on the last day of each month. The matching indices are the last month's index and this month's index. For example, if the monetary policy is published on March 10th, then the indices released on Feb 28th and March 31st are considered matching for this policy. The trend is calculated by the differential of the two indices. If the newer one is bigger than the previous one, the trend is "up", otherwise it's "down". That means zero change is considered down.

When doing this process, the author found a few problems, and here are the solutions took:

The first monetary decision was published on March 4th, 1999 while the second was published on March 18th, 1999, which means there is no matching index for the first one. The first one was removed to fix this problem.

Three monetary decisions were published on April 30th, 2020. These three decisions were merged and considered a single one published on that day.

After these steps, the number of available texts decreased from 293 to 290. Therefore, 290 analysis objects were obtained and ready for further analysis.

4. Methodology

4.1. A brief introduction to the models used in the study

In this study, we aimed to predict the trend of two indices following each monetary policy decision using nine typical classification models provided by scikit-learn. The problem at hand is a simple binary classification task, and after obtaining the individual results from each model, we applied an ensemble method to merge these results. The ensemble method combines the predictions of the nine models and outputs either an increase or decrease prediction based on the majority vote.

Subsequently, we compiled a table with all the results and calculated the accuracy, precision, recall, and F-score to evaluate the performance of the model. The nine models utilized in this study and their evaluation metrics are explained in detail below.

4.1.1. K-neighbors classifier. K- nearest neighbor algorithm is a kind of classification algorithm, which belongs to lazy learning [36]. If most of the k nearest samples of a sample in the feature space belong to a certain category, the sample is also classified into this category. The value of K is very important. The common method is to estimate the error rate of the classifier by using the test set starting from k=1. Repeat the process, each time K increases in value by 1, allowing for the addition of one neighbor. Choose K with the lowest error rate.

Euclidean distance is used to calculate the distance between two points. The definition of the Euclidean distance is, given $p_1(x_1, y_1), p_2(x_2, y_2)$, where p_i represents the point of location, the formula is $\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$. Hence, for the multi - dimension, the formula is $E(x, y) = \sqrt{\sum_{i=0}^n (x_i - y_i)^2}$. Consequently, comparing the distance of each point to get the K - nearest neighbor, and put the sample into a certain category.

4.1.2. Support vector classifier. Support vector classifier (SVC) is a binary classification model whose basic model is a linear classifier with the largest interval defined on the feature space [37]. Suppose that the training data set of the feature space is $T = (x_1, y_1), (x_2, y_2) \dots (x_{n-1}, y_{n-1}), (x_n, y_n)$. Simply speaking, looking for a hyperplane that divides a training data set into two categories, and the formula for finding the hyperplane is $w * b + x = y$, then $|w * x + b|$ could be the distance from x to hyperplane, $\gamma = y/(|w * x + b|)$ could judge the correctness of classification. Since function interval endorsement does not change the hyperplane, it can take $\gamma = 1$, then the constriction could be $y/(|w * x + b|) \geq 1$. In general, linearly indivisible means that the function margin cannot bigger or equal to 1. Hence, a slack variable should be introduced, and the formula should change to $y/(|w * x + b|) \geq 1 - \xi$ Finally, the SVM model could change to a convex quadratic optimization problem.

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i(w * x_i + b) \geq 1 - \xi_i \quad i = 1, 2, \dots, n \\ \text{s.t.} \quad & \xi \geq 0, \quad i = 1, 2, \dots, n \end{aligned}$$

where w is the weight vector, b is the bias term, x_i is the feature vector for the i-th training example y_i is the class label for the i-th training example (+1 or -1), C is the regularization parameter that controls the trade-off between maximizing the margin and minimizing the classification error.

4.1.3. Gaussian process classifier. The Gaussian Process Classifier is a probabilistic classification algorithm [38]. It uses Gaussian processes to model the decision boundary between classes. The algorithm models the relationship between the input features and the output classes as a probability distribution over functions. Firstly, the algorithm works by a prior distribution over functions. The next step is updating this distribution based on the observed data. Finally, the algorithm makes predictions using the posterior distribution for new inputs.

4.1.4. Decision tree classifier. A decision tree classifier is an attribute structure where each internal node represents a test on an attribute, each branch represents a test output, and each leaf node represents a category [39]. The decision tree is a case - based inductive learning which adopts the recursive method from top to bottom. The classification criteria of decision tree can choose Gini coefficient or information entropy. The entropy of 1 represents the most disorderly state, and the entropy of 0 represents that the leaf nodes of the decision tree belong to the same category. When all the data in a subset belongs to the same class or satisfies the criterion, the classification will be done.

4.1.5. Naive bayes classification. Naive Bayes classifier is based on Bayes algorithm, the core formula is $P(B|A) = \frac{P(A|B)P(B)}{P(A)}$ In the naive Bayes algorithm, this equation is true if the characteristic conditions are independent of each other [40]. Then for n different factors, the formula will change to $P(B|A_1, A_2, \dots, A_n)$, then the conditional distribution for B could be expressed as $P(B|A_1, A_2, \dots, A_n) =$

$P(A_1, A_2, \dots, A_n|B) * P(B)/P(A_1, A_2, \dots, A_n)$, where $P(B|A_1, A_2, \dots, A_n)$ represents the probability of class B given feature value A_1, A_2, \dots, A_n . $P(A_1, A_2, \dots, A_n|B)$ is the probability of feature values A_1, A_2, \dots, A_n given class B, $P(A_1, A_2, \dots, A_n)$ is the marginal probability of (A_1, A_2, \dots, A_n) . $P(y)$ is the prior probability of class y, which is calculated from the frequency of each class in the training set. The prior probability of B is: $p(B = b_1) = \frac{1}{\text{the number of sample}}$

And the conditional probability of B : $p(A_i = a_i|B = b_1) = \frac{\text{the number of sample } A_i=a_i}{\text{the number of total sample}}$

Then the conditional probability of the class can be obtained. Finally, the probability values of each class can be compared to determine which class the test sample should belong to.

4.1.6. Quadratic discriminant analysis. Quadratic discriminant analysis (QDA) is a classification algorithm that uses a quadratic decision boundary to estimate the probability that a point belongs to a particular class [41]. The algorithm assumes that the data is generated from a multivariate Gaussian distribution and estimates each class's mean and covariance matrix. Given a new data point, the algorithm uses Bayes' theorem (talk about in the previous method) to calculate the posterior probability of data belonging to each class. The class with the highest probability is then assigned to the data point. The decision boundary in QDA is a quadratic function that allows for more flexibility than linear decision boundaries in cases where classes are not well separated.

4.1.7. Adaboost classifier. AdaBoost (Adaptive Boosting) Classifier is a boosting ensemble algorithm that combines several weak classifiers to create one strong classifier [42]. The algorithm sequentially adds weak learners to the model, each trained on data where previous learners made mistakes. The final model consists of a weighted sum of weak learners and assigns weights based on their performance. And the formula could be summarized as $w_1 = \frac{1}{N}$, w_1 means the initial observation weight, and the N means the number of the training sample.

For each weak learner:

- 1) Training a weak learner on the training data using the current weights.
- 2) Compute the error rate of the weak learner.
- 3) Compute the weight of the weak learner alpha.
- 4) Update the weights of the training data based on whether the weak learner got each sample correctly or incorrectly.

Finally, weak learners are combined into strong learners by weighting their alpha values. Then the strong learner could make predictions.

4.1.8. MLP classifier. The Multi-layer Perceptron Classifier (MLP Classifier) is an artificial neural network that uses multiple layers to classify data. Its algorithm principle is based on the idea of feedforward artificial neural network, where data is fed forward through multiple layers of nodes, and each layer performs some calculations on the data before passing it on to the next layer [43]. Between the input layer and the output layer is called the hidden layer, where the node section connects the previous layer to the subsequent layer. The output formula of the hidden node is: $Z = w_1x_1 + w_2x_2 + w_3x_3 + \dots + w_nx_n + b$, where w_i are the weights of the connections between the inputs (x_i) and the node, b is the bias term, and z is the weighted sum of the inputs. This value is then passed through an activation function and produces the output of the node. Then, repeat the process to get the final answer.

4.1.9. The random forest classifier. Random forest classifiers use multiple decision trees for prediction. It works by creating a set of decision trees from a randomly selected subset of training data [44]. During the creation process, the algorithm also randomly selects a subset of features to consider at each split. After the build is complete, the predictions are made by majority voting on the predictions of a single number. Each decision tree is predicted using the following formula:

$$\text{PredictedClass} = \text{argmax}(P(Y = 1|X), P(Y = 2|X) \dots P(Y = k|X))$$

where Y is the target variable, X is the feature vector, k is the number of classes, and argmax is a function that returns the class with the highest probability. The probability of each class is from the leaf node of decision trees.

4.2. The utilization of the models. The TF-IDF table was divided into two parts, with 80% of the data used for training the classification models and the remaining 20% used for model evaluation. The models were designed to predict the change in indices, with a "1" indicating an increase and a "0" indicating a decrease.

After the models were developed, a grid search was conducted to identify the optimal parameters for each model. The best parameters were recorded in a text file and utilized during the model evaluation phase. The results of each model were recorded in a table and used for post-processing. The true positive, false negative, false positive, and successful predictions were calculated during this stage. These metrics were used to generate the final evaluation table.

In summary, the methodology involved splitting the data into training and evaluation sets, using classification models to predict changes in indices, and post-processing the results to evaluate the model's performance.

5. Result

The texts involved in this study are highly official. Unlike the lived texts presented in the references, official documents do not tend to contain strong emotional overtones, which poses a challenge for the analysis. Conventional preprocessing and fitting converge to zero, so a combination of ensemble and multiple machine learning schemes is chosen to ensure that predictions are made with as high a precision as possible for a small amount of text.

Several experiments have been completed to verify the model's reliability, and their optimal results have been recorded. The tests were based on the following specifications of the equipment:

Macbook Pro 2021 M1 Max with 32-core GPU 64G RAM

For the software part, Visual Studio Code and Python 3.11.2 ARM64 was used.

The time used for each model during the experiment has been recorded and shown in Table 2.

Table 2. The time consumed for each model to train and infer.

Model	Training Time(in seconds)	Inference Time(in seconds)
KNeighborsClassifier	0.010515928	0.017198086
SupportVectorClassifier	0.041743755	0.007359028
GaussianProcessClassifier	49.46303606	0.040885925
DecisionTreeClassifier	0.020855665	0.003522158
RandomForestClassifier	0.134315014	0.023402929
MLPClassifier	0.446545839	0.003533125
AdaBoostClassifier	1.516850233	0.051281929
GaussianNB	0.008069038	0.003450871
QuadraticDiscriminantAnalysis	0.09889102	0.004673004

Using a set of result retrieval procedures, we found the best results from the training of each model.

Finally, we compared the actual indices with our predicted results, and the results of the comparison are shown in Table 3.

Table 3. The evaluation table for each model.

Model	Accuracy	Precision	Recall	F-score
KNeighborsClassifier	0.965517241	0.939393939	1	0.96875
SupportVectorClassifier	0.965517241	0.939393939	1	0.96875
GaussianProcessClassifier	0.982758621	0.96875	1	0.984126984
DecisionTreeClassifier	0.965517241	0.967741935	0.967741935	0.967741935
RandomForestClassifier	0.948275862	0.9375	0.967741935	0.952380952
MLPClassifier	0.948275862	0.911764706	1	0.953846154
AdaBoostClassifier	0.965517241	0.939393939	1	0.96875
GaussianNB	0.827586207	0.769230769	0.967741935	0.857142857
QuadraticDiscriminantAnalysis	0.879310345	0.833333333	0.967741935	0.895522388
Ensemble	0.965517241	0.939393939	1	0.96875

Overall, by combining multiple machine learning methods, results arrive at predicted values that are more similar to the actual values. Therefore, it can be said that this prediction algorithm based on NLP and mechanical learning techniques is relatively effective in predicting the future ESI and EEI indices in Europe.

6. Discussion

In this study, we implemented several classic classification models to predict the trend of EEI and ESI after each monetary policy decision. To determine the effective indices of each decision, we selected the indices released 30 days after each decision and the most recent indices released before the decision. The difference between the two indices was used to determine the trend. However, monetary policy decisions are not always synchronized with index releases, as decisions are typically published every 2-3 months while indices are released monthly. To address this issue, we extended the effective period of each decision to include the time between the previous and the next decision. Any indices falling within this period were considered influenced by the decision, and trends were calculated as before.

The results showed that only 8 out of 288 texts differed between the trend calculated using the original 30-day period and the extended effective period, which suggests that predicting trends based on 30 days is reasonable.

One notable characteristic of the results was that some models had a full recall, which may be due to the nature of the indices, which typically increase over time and only drop in dire situations such as the financial crisis, COVID-19 outbreak, and the Ukraine conflict. Since positive results (i.e., an increase in indices) were defined as "1" in this study, false negatives were less likely to occur. Overall, the results suggest that the models effectively predict indices and can be used confidently in production.

7. Conclusion

This paper investigates implementing NLP techniques in forecasting the trend of two indexes from "The Business and Consumer Surveys indicators", the EEI and ESI, to reflect the consumers' reaction to capture the consumers' response. The performance of the NLP model is evaluated based on its accuracy, precision, recall, and F1 score in forecasting the actual outcomes.

By analyzing results from 9 models by sci-kit-learn, the highest model could reach 98% in accuracy and F-recall, the lowest model still being able to achieve 82% in accuracy. The results of this study indicate that the model is effective in forecasting. NLP techniques are an effective approach for predicting the trends of these economic indicators, and the model developed exhibits a high level of performance. These findings provide valuable insights for policymakers and stakeholders in the economic sector, as they can use this approach to make informed decisions based on consumer behaviour trends. Unfortunately, the result of the forecasting is closer to the test of the model's performance and

less relative to the forecasting of the expected value of the index in the future. Our data set was not big enough to support further step research.

Future research could consider more data sets from multiple fields in forecasting. This approach would involve identifying and including as many factors as possible that significantly impact ESI and EEI to minimize the effect of the error term in the research. By incorporating a broader range of factors that affect ESI and EEI, such as macroeconomic indicators, industry-specific data, and labour market trends, it can develop more robust models for forecasting these indicators.

References

- [1] Susan V. Scott & Markos Zachariadis (2012) Origins and development of SWIFT, 1973–2009, *Business History*, 54:3, 462-482, DOI: 10.1080/00076791.2011.638502
- [2] Lo, Gaye-Del & Marcelin, Isaac & Bassène, Théophile & Sène, Babacar (2022) "The Russo-Ukrainian war and financial markets: the role of dependence on Russian commodities," *Finance Research Letters*, Elsevier, vol. 50(C).
- [3] Frederic S. Mishkin (2016) *The Economics of Money, Banking, and Financial Markets* (Eleventh Edition), Pearson.
- [4] Lo, G. D., Marcelin, I., Bassène, T., & Sène, B. (2022, December). The Russo-Ukrainian war and financial markets: the role of dependence on Russian commodities. *Finance Research Letters*, 50, 103194. <https://doi.org/10.1016/j.frl.2022.103194>
- [5] The Fed. (2022, June 21). The Fed - Monetary Policy: Monetary Policy Report. The Fed - Monetary Policy: Monetary Policy Report. Retrieved November 15, 2022, from <https://www.federalreserve.gov/monetarypolicy/2022-06-mpr-summary.htm>
- [6] deLisle, J. (2022). Deterrence Dilemmas and Alliance Dynamics: United States Policy on Cross-Strait Issues and the Implications of the War in Ukraine. *American Journal of Chinese Studies*, 29(2).
- [7] Lehmann, R. The Forecasting Power of the ifo Business Survey. *J Bus Cycle Res* (2022). <https://doi.org/10.1007/s41549-022-00079-5>
- [8] Hoekstra, Janny C., and Peter SH Leeftang. "Thriving through turbulence: Lessons from marketing academia and marketing practice." *European Management Journal* (2022).
- [9] Lo, G. D., Marcelin, I., Bassène, T., & Sène, B. (2022, December). The Russo-Ukrainian war and financial markets: the role of dependence on Russian commodities. *Finance Research Letters*, 50, 103194. <https://doi.org/10.1016/j.frl.2022.103194>
- [10] Zhang, T., Yang, K., Ji, S., & Ananiadou, S. (2023). Emotion fusion for mental illness detection from social media: A survey. *Information Fusion*, 92, 231-246.
- [11] Cui, Y., Jiang, Y., & Gu, H. (2023, January). Novel Sentiment Analysis from Twitter for Stock Change Prediction. In *Data Mining and Big Data: 7th International Conference, DMBD 2022, Beijing, China, November 21–24, 2022, Proceedings, Part II* (pp. 160-172). Singapore: Springer Nature Singapore.
- [12] Alsiaity, A., & Orji, R. (2022). Machine learning techniques for emotion detection and sentiment analysis: current state, challenges, and future directions. *Behaviour & Information Technology*, 1-26.
- [13] Ruby, D., & About The Author Daniel Ruby Content writer with 10+ years of experience. I write across a range of subjects. (2023, March 1). 58+ twitter statistics for marketers in 2023 (Users & Trends). Demand Sage. Retrieved March 18, 2023, from <https://www.demandsage.com/twitter-statistics/#:~:text=Twitter%20has%20around%20450%20million%20monthly%20active%20users%20as%20of%202023>
- [14] Tiwari, A. K., Abakah, E. J. A., Bonsu, C. O., Karikari, N. K., & Hammoudeh, S. (2022). The effects of public sentiments and feelings on stock market behavior: Evidence from Australia. *Journal of Economic Behavior & Organization*, 193, 443-472.

- [15] Sawicki, G. S., Chilvers, M., McNamara, J., Naehrlich, L., Saunders, C., Sermet-Gaudelus, I., ... & Davies, J. C. (2022). A Phase 3, open-label, 96-week trial to study the safety, tolerability, and efficacy of tezacaftor/ivacaftor in children ≥ 6 years of age homozygous for F508del or heterozygous for F508del and a residual function CFTR variant. *Journal of Cystic Fibrosis*, 21(4), 675-683.
- [16] Guzmán, R., & Morales, G. Discursive Strategies and Assessment in Turing test: A developmental analysis of L2 acquisition.
- [17] Fan, J., Campbell, M., & Kingsbury, B. (2011). Artificial intelligence research at IBM. *IBM Journal of Research and Development*, 55(5), 16-1.
- [18] Torfi, A., Shirvani, R. A., Keneshloo, Y., Tavaf, N., & Fox, E. A. (2020). Natural language processing advancements by deep learning: A survey. *arXiv preprint arXiv:2003.01200*.
- [19] Chen, S., Wang, C., Chen, Z., Wu, Y., Liu, S., Chen, Z., ... & Wei, F. (2022). Wavlm: Large-scale self-supervised pre-training for full stack speech processing. *IEEE Journal of Selected Topics in Signal Processing*, 16(6), 1505-1518.
- [20] Friederich, S. (2017). Fine-tuning.
- [21] Jolicoeur, P., & Jolicoeur, P. (1999). Fisher's linear discriminant function. *Introduction to biometry*, 303-308.
- [22] Naïve Bayes Classifiers Revisited, Domingos, P., 1997.
- [23] A Few Useful Things to Know About Machine Learning, Pedro Domingos, 2012.
- [24] Gradient-based learning applied to document recognition, LeCun, Y. et al., 1998.
- [25] Long Short-Term Memory, Hochreiter, S. & Schmidhuber, J., 1997.
- [26] Random Forests, Breiman, L., 2001.
- [27] A Logical Calculus of the Ideas Immanent in Nervous Activity, McCulloch, W. S. & Pitts, W., 1943.
- [28] The perceptron: a probabilistic model for information storage and organization in the brain, Rosenblatt, F., 1958.
- [29] Some studies in machine learning using the game of checkers, Samuel, A. L., 1962.
- [30] Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences, Werbos, P. J., 1975.
- [31] Learning internal representations by error propagation, Rumelhart, D. E., Hinton, G. E. & Williams, R. J., 1986.
- [32] A Tutorial on Support Vector Machines for Pattern Recognition, Burges, C. J. C., 1998.
- [33] European Central Bank. (2020, January 8). ECB Monetary policy decisions. European Central Bank. Retrieved March 18, 2023, from <https://www.ecb.europa.eu/press/govcdec/mopo/html/index.en.html>
- [34] European Central Bank. (n.d.). ECB Latest Business and Consumer Surveys. Economy and Finance. Retrieved March 18, 2023, from https://economy-finance.ec.europa.eu/economic-forecast-and-surveys/business-and-consumer-surveys/latest-business-and-consumer-surveys_en
- [35] Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513-523.
- [36] A method of comparing two groups of learning, Cover, T. & Hart, P., 1967
- [37] Pu, Y., Apel, D. B., & Xu, H. (2019). Rockburst prediction in kimberlite with unsupervised learning method and support vector classifier. *Tunnelling and Underground Space Technology*, 90, 12-18.
- [38] Gibbs, M. N., & MacKay, D. J. (2000). Variational Gaussian process classifiers. *IEEE Transactions on Neural Networks*, 11(6), 1458-1464.
- [39] Du, W., & Zhan, Z. (2002). Building decision tree classifier on private data.
- [40] Rish, I. (2001, August). An empirical study of the naive Bayes classifier. In *IJCAI 2001 workshop on empirical methods in artificial intelligence* (Vol. 3, No. 22, pp. 41-46).

- [41] Ghojogh, B., & Crowley, M. (2019). Linear and quadratic discriminant analysis: Tutorial. arXiv preprint arXiv:1906.02590.
- [42] Rojas, R. (2009). AdaBoost and the super bowl of classifiers a tutorial introduction to adaptive boosting. Freie University, Berlin, Tech. Rep.
- [43] Hampshire II, J. B., & Pearlmutter, B. (1991). Equivalence proofs for multi-layer perceptron classifiers and the Bayesian discriminant function. In *Connectionist Models* (pp. 159-172). Morgan Kaufmann.
- [44] Kulkarni, V. Y., & Sinha, P. K. (2012, July). Pruning of random forest classifiers: A survey and future directions. In *2012 International Conference on Data Science & Engineering (ICDSE)* (pp. 64-68). IEEE.