

NoB_1 Data Cleaning and Content Discoveries

1. Data ingestion and Initial inspection

1.1 Introduction

1.2 Key information from the Metadata

- I. Data dictionary
- II. Original and Final Data after cleaning

1.3 Setting up the Notebook workflow

- I. Imports
- II. Configuration settings and constants
- III. Functions to load and save data

1.4 Sense Checking Raw Data

- *Critical: Economic Indicator selection and rationale*

2. Data cleaning and Text data Preprocessing

2.1 Data Cleaning

2.2 Sense Checking resulting DataFrames

2.3 Text data Preprocessing

- *Objective: Clean DataFrames, saved them as .csv files*

3. Exploratory Analysis and Context discoveries

3.1 Exploratory Data Analysis (EDA)

- I. Data Distribution Visualisations
- II. Summary Statistics for economic indicators
- III. Outliers and Distribution

3.2 Time Series Visualisations

- *Objective: Economic Trends and Patterns*

4. Sentiment Analysis on Speech Data

- 4.1 Naive Bayes Sentiment Classifier
 - I. WordClouds and Frequency Distribution
 - II. Polarity and Sentiment Score
 - III. Top 20 Positive and Negative Speeches
 - 4.2 Loughran and McDonald Sentiment Word List
 - 4.3 Exploratory Data Visualisations
 - I. Timeline and Speech Sentiment Changes
 - II. Pattern Investigation
- *Objective: Patterns and trends*

NoB_2 FinBert and Correlation Analysis: OLS and VAR

1. Data ingestion and Inspection

- 1.1 Introduction
- 1.2 Key information from the Metadata
 - I. Original and Final Data after Analysis
- 1.3 Setting up the Notebook workflow
 - I. Imports
 - II. Configuration settings and constants
 - III. Functions to load and save data
- 1.4 Sense Checking the DataFrames

2. FinBert Sentiment Analysis

- 2.1 FinBert Sentiment Anlysis
 - I. 1st Attempt and visualisations (512 Tokens Truncation)
 - II. 2nd Attempt and visualisations (FinBert by Statements)
 - III. Visualisations

- *Objective:*
Generate a sentiment score classification (final df creates inbert_speeches.csv).

3. Correlation Analysis: OLS Regression

- 3.0 DataFrames Merge
- 3.1 Simple Linear Regression
 - I. Correlation Matrix
 - II. Simple Linear Regression for each of the variables
- 3.2 Multiple Linear Regression
 - I. Variance-inflation Factor (VIF)
 - II. Multiple Linear Regression

- *Objective: Identify correlations between sentiment and economic indicators.*

4. Correlation Analysis: Vector Autoregression (VAR) Model

4.1 Vector Autoregression (VAR) Model

- I. Data Preparation
- II. Model Fit and Visualisations

- *Objective: Analyse dynamic interactions between sentiment and economic indicators.*

NoB_3 Finbert and Predictive Models: Random Forest and SARIMA

1. Data ingestion and Inspection

- 1.1 Introduction
- 1.2 Key information from the Metadata
 - II. Original and Final Data after Analysis
- 1.3 Setting up the Notebook workflow
 - IV. Imports
 - V. Configuration settings and constants
 - VI. Functions to load and save data
- 1.4 Sense Checking the DataFrames

2. Predictive Modeling: Random Forest

- 2.0 Data Preparation
- 2.1 Random Forest
 - I. Correlation matrix and VIF
 - II. Random Forest Model Fit and Feature Importances

3. Predictive Modeling: SARIMA Model

- 3.1 Sesonal AutoRegressive Integrated Moving Average (SARIMA) Model
 - I. Data Preparation
 - II. SARIMA Model Fit
 - III. Residual diagnostics and Statistical test
 - IV. Forecast and Visualisations
- *Objective: Predictive model for forecasting the tone of upcoming Bank of England speeches based on past economic indicators.*