***Inflation Rate Classification Based on Economic Indicators***

**1 PROJECT OVERVIEW**

**Background**

This project investigates whether public macroeconomic data can be utilized to create an inflation thermometer. We’re tracking monthly data, including Headline and Core CPI, final-demand PPI, unemployment rates, industrial production, Treasury produce spreads, and short-term inflation expectations (Baybuza, 2018; Jin and Simionescu, 2025; Khashimova and Buranova, 2023). All of this data comes from the Federal Reserve’s FRED portal and covers the years 1970 to 2024.

From these data sets, we obtain several essential features, including levels, year-over-year changes, volatility, and a flag for recessions. We categorize inflation into three functional groups: low (under 2%), moderate (2–5%), and high (over 5%).

We will first label each monthly inflation observation by applying a simple rule-based function, classify inflation which assigns 'Low' when the rate is below 2 percent, 'Moderate' when it is at least 2 percent but below 5 percent, and 'High' when it reaches 5 percent or more. The resulting Inflation Category column will serve as the target variable for our machine-learning pipelines, enabling both the SVM and Neural Network to learn how these three regimes relate to the underlying economic indicators.

**Research Questions**

* Can we use economic indicators (like CPI, PPI, and unemployment rate) to classify inflation periods?
* What combinations of features give the best classification results?
* How well do models like SVM and Neural Networks deal with seasonal changes in the economy?

**Project Objectives**

* To build a cleaned and merged monthly dataset of CPI, PPI, unemployment rates, and other relevant variables from FRED.
* To create features based on lags, growth rates, and volatility, then choose the most useful ones.
* To train and fine-tune SVM and shallow Neural Network classifiers, comparing their performance over time.
* To test how sensitive the models are to seasonality by running tests on both seasonally adjusted and raw data.

References

Baybuza, I., 2018. Inflation Forecasting Using Machine Learning Methods. Russian Journal of Money and Finance 42–59. https://doi.org/10.31477/RJMF.201804.42

Jin, B., Simionescu, M., 2025. Machine Learning vs. Econometric Models to Forecast Inflation Rate in Romania? The Role of Sentiment Analysis. Mathematics 2025, Vol. 13, Page 168 13, 168. <https://doi.org/10.3390/MATH13010168>

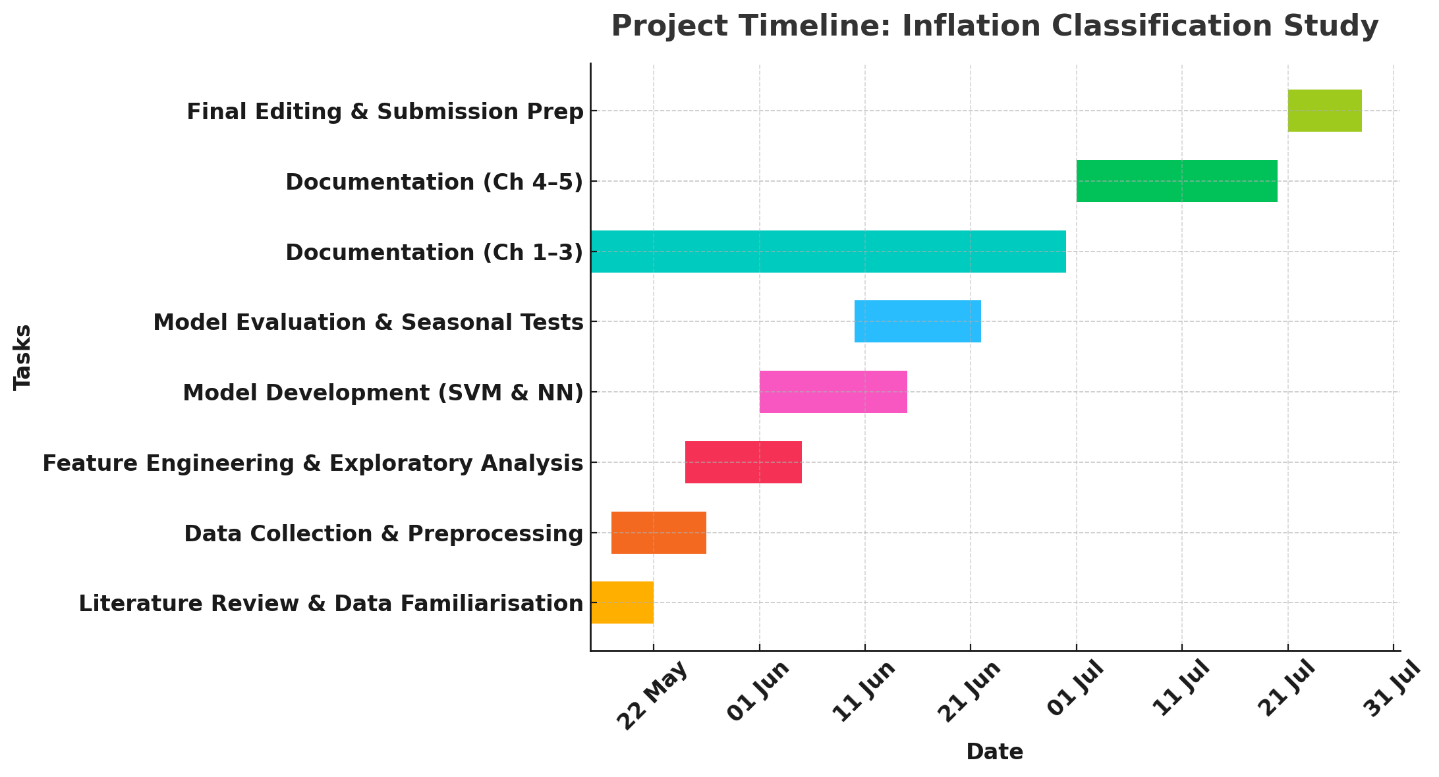
Khashimova, N., Buranova, M., 2023. Comparative Analysis of Machine Learning Algorithms for Inflation Rate Classification and Economic Trend Forecasting. ACM International Conference Proceeding Series 274–282. https://doi.org/10.1145/3644713.3644749

***2 PROJECT PLAN: TASK LIST AND TIMELINE***

**Task Dates Description**

|  |  |  |
| --- | --- | --- |
| **Task** | **Dates** | **Description** |
| Literature review & data Exploration | 16 May – 22 May | Read recent papers and explore FRED series definitions. |
| Data collection & preprocessing | 18 May – 27 May | Download monthly CPI, PPI, unemployment, and metadata; handle missing values; align frequencies. |
| Feature engineering & exploratory analysis | 25 May – 5 June | Create growth and volatility features, visualize class boundaries, and rank variables. |
| Model development (SVM & NN) | 1 June – 15 June | Grid-search hyperparameters, implement cross-validation pipelines. |
| Model evaluation & seasonal robustness tests | 10 June – 22 June | Compare seasonally adjusted vs raw inputs under rolling windows. |
| Documentation (Ch 1–3) | 16 May – 30 June | Write an introduction, literature review, and methodology while the experiments run. |
| Documentation (Ch 4–5) | 1 July – 20 July | Draft results, discussion, and conclusions. |
| Final editing & submission preparation | 21 July – 28 July | Proofread, format, conduct a final Turnitin check, and upload to Canvas. |

**Gantt Chart**



***3 DATA MANAGEMENT PLAN***

**Overview of the Dataset**

We’ll use monthly economic indicators from the United States, provided by the Federal Reserve Economic Data platform and gathered initially by the Bureau of Labor Statistics and other related agencies. There’s no personal data involved here.

**Data Collection**

We’ll write Python scripts using the pandas-datareader FRED API to download CPI, PPI, the unemployment rate, and some other series into a data/raw folder. We’ll document every step to ensure it can be easily reproduced.

**Metadata**

Raw CSV files will include series codes, observation dates, and values, with each file staying under 1 MB. After processing, we expect the resulting dataset to be approximately 2 MB in size, with model checkpoints and logs adding 50 MB.

**Document Control**

All code and data will be stored in a GitHub repository at github.com/&lt;user&gt;/inflation-classification. We’ll push updates weekly and review pull requests to ensure quality maintenance.

**ReadMe File**

In the final ReadMe, we will summarize the project aims, data sources, and setup instructions, and guide reproducing tables and figures, along with contact information for further collaboration.

**Security and Storage**

All raw data and code will be kept in GitHub and a mirrored OneDrive folder. GitHub will serve as the primary version control system, while OneDrive provides daily backups. Sensitive information will be excluded using .gitignore.

**Ethical Requirements**

**GDPR:** We only use aggregated data that can’t identify individuals, so GDPR and UK DPA rules don’t apply.

**UH Ethics:** Public secondary data that isn’t personal doesn’t need a full review, but we’ll keep things clear by citing sources, sharing our code, and providing a replication appendix.

**Licensing &amp; Attribution:** FRED allows non-commercial academic use if you give credit; all the U.S. government data we use is public and will be acknowledged appropriately.

**Provenance:** The data originates from trusted government sources that employ approved statistical methods, ensuring that there are no ethical concerns regarding its collection or potential biases.

**Data Stewardship:** Raw data is only accessible in a read-only format. Processed data and code are kept in versions on GitHub, with backups on OneDrive to ensure security and prevent accidental leaks.