# Project Proposal: Predicting Future Military Expenditures Using SIPRI Data

Project Title: Forecasting Military Expenditure and Analyzing Spending Patterns Across Nations

## Project Overview and Problem Definition

### Objective

The goal of this project is to predict future military expenditures for various countries based on historical data from the **Stockholm International Peace Research Institute (SIPRI)** and related economic and geopolitical factors. By leveraging SIPRI’s datasets on military spending, arms transfers, and conflict, we aim to build a time series forecasting model that can provide insights into spending trends by country and region. Additionally, we will analyze key drivers of these expenditures, such as GDP and regional conflict intensity to identify patterns in global military spending.

### System Overview

The system we aim to build will consist of a forecasting model that predicts annual military expenditure for each country based on historical spending data and additional socio-economic indicators. A clustering component will then group countries with similar spending patterns, offering a high-level understanding of regional shifts in military expenditure.

## Algorithms and Techniques to Investigate

Probabilistic Reasoning over Time (Chapter 14)

Adapt probabilistic models to analyze temporal patterns in military spending.

Supervised Learning (Regression, Chapter 19)

Predict future military expenditure using features like GDP, arms imports, etc.

Unsupervised Learning (Clustering, Chapter 20)

Group countries with similar spending patterns using clustering algorithms such as K-Means.

Feature Selection and Dimensionality Reduction (Chapter 19)

Techniques to optimize features impacting predictions.

## Related Course Topics

The following topics are directly covered in the course textbook and align with our project focus:

Probabilistic Reasoning over Time (Chapter 14)

Enables analysis of spending trends and temporal dependencies.

Supervised Learning (Regression, Chapter 19)

Allows for accurate prediction of future spending based on historical data.

Unsupervised Learning (Clustering, Chapter 20)

Useful for grouping countries by spending patterns, providing insights into regional trends.

Feature Selection and Dimensionality Reduction (Chapter 19)

Helps in identifying key predictors of military expenditure.

## Expected System Behavior

The system is expected to:

Accurately predict future military expenditure for individual countries over the next 5-10 years.

Identify the main socio-economic and political drivers of changes in spending patterns.

Group countries into clusters based on similarities in expenditure trends or geopolitical factors.

Examples of expected behavior include:

Predicting Expenditure Trends: Producing an annual expenditure forecast for selected countries.

Driver Analysis: Highlighting which factors (e.g., GDP, regional conflict) are most influential.

Clustering for Regional Analysis: Visualizing clusters of countries with similar trends, revealing potential geopolitical patterns.

## Key Issues and Challenges

The primary challenges anticipated include:

**Data Variability** - Military expenditure data varies significantly across countries, introducing high variability and requiring normalization.

**Feature Selection** - Identifying and selecting impactful socio-economic and political features that drive expenditure changes.

**Model Complexity and Interpretability** - Balancing model complexity for high accuracy with interpretability, especially if using advanced models.

## Resources and References

The following resources will be foundational to our project research and development:

### SIPRI Databases:

SIPRI Military Expenditure Database

Contains military expenditure data by country from 1949.

[https://www.sipri.org/sites/default/files/SIPRI-Milex-data-1948-2023.xlsx](https://www.sipri.org/sites/default/files/SIPRI-Milex-data-1948-2023.xlsx%20)

### Other Possible Datasets:

* **World Bank Indicators**

World Bank indicators for GDP and political stability

<https://datacatalog.worldbank.org/search/dataset/0037712/World-Development-Indicators>

* **ACLED Conflict Index**

Ranking violent conflict levels across the world, 2024

### <https://acleddata.com/conflict-index/>

### Time Series Forecasting Research:

Hyndman, R.J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. OTexts.

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. Neural Computation.

### Clustering and Geopolitical Analysis:

Kaufman, L., & Rousseeuw, P.J. (2009). Finding Groups in Data: An Introduction to Cluster Analysis. Wiley.

# Team Contributions

Project Assignments are in flux at this early stage but research and starting goals have been assigned per below for the initial phase.

|  |  |  |  |
| --- | --- | --- | --- |
| Task | Description | Week | Assigned |
| Communication | Slack Channel Created | 2 | Swapnil |
| Zoom Meetings | Zoom Meeting Invites (Weekly, Tues/Fri) | 2 | Swapnil |
| Project Proposal |  | 3 | Carrie |
| Git Repo and Readme | <https://github.com/swapnilprakashpatil/aai501_6proj> | 3 | Swapnil |
| Coding | Data Analysis and Cleaning/Preprocessing | 3, 4, 5 | Carrie |
|  | Algorithm Selection |  |  |
|  | Algorithm Analysis |  |  |
|  | Implementation |  |  |
|  | Probabilistic Reasoning over Time | 3, 4, 5 | Maurine |
|  | Supervised Learning (Regression) | 3, 4, 5 | Carrie |
|  | Unsupervised Learning (Clustering) | 3, 4, 5 | Swapnil |
|  | Feature Selection and Dimensionality Reduction | 3, 4, 5 | ALL |
|  | Hyperparameter Tuning |  |  |
|  | Algorithm output and Metrics |  |  |
|  | Comparative Analysis, Evaluation and Critique |  |  |
|  | Hypothesis testing and validation |  |  |
|  | Code Review |  |  |
| Final Paper | Purpose, Goal and Scope | 6 | TBD |
|  | Data Analysis and Preprocessing |  |  |
|  | Algorithm Selection |  |  |
|  | Algorithm Analysis |  |  |
|  | Algorithm output and Metrics |  |  |
|  | Comparative Analysis |  |  |
|  | Conclusion |  |  |
|  | Limitations and Next Steps |  |  |
|  | Appendix with collaboration details |  |  |
|  | References |  |  |
|  | Turnitin Similarity Fixes |  |  |
| Presentation & Video | Overview | 7 | TBD |
|  | Business Use cases |  |  |