**PROPOSAL**

We propose to analyze Walmart's weekly sales dataset, which contains data on sales revenue, store location, date, and various environmental factors such as temperature, fuel price, CPI, and unemployment. Our objective is to develop a predictive model to forecast weekly sales based on these factors. The analysis will help Walmart to optimize its inventory management, pricing, and marketing strategies and improve overall business performance. By enhancing inventory control, cutting waste, and increasing the precision of sales forecasts, it may be possible to save money. However, there are also dangers in employing AI solutions, such as the necessity to safeguard sensitive data and the possibility of bias in the algorithms.

**Methodology**: Our proposed methodology involves the following steps:

**Data cleaning and preparation**: We will remove any missing or inconsistent data and convert categorical variables to numerical formats.

**Exploratory data analysis**: We will conduct an exploratory analysis of the data to identify any patterns or correlations between the various variables.

**Feature engineering**: We will engineer additional features from the existing data.

**Model selection and evaluation:** We will train and evaluate several machine learning models, such as linear regression, decision trees, random forests, Support Vector Regression and neural networks. We will use appropriate evaluation metrics such as mean absolute error and mean squared error to compare the performance of these models.

Accuracy of linear regression model is: 0.15.

Accuracy of Decision Tree Regression model is: 0.88.

Accuracy of Support Vector Regression-0.028

Accuracy of Random Forest model is 0.93.

Accuracy of Neural network model is: -2.77

This variation may be due to Model Complexity, overfitting. For our research, we tested five number of models and discovered that the Random Forest model had the best accuracy score (0.93), followed by the Decision Tree Regression model (0.88). The accuracy score for the linear regression model was 0.15, which was much less accurate than the scores for the other models. Poor performance was shown by the Support Vector Regression model's accuracy score of -0.028 and the Neural Network model's accuracy score of -2.77.

We will utilise the Random Forest model for our project since it can handle complicated interactions between characteristics in the dataset and got the highest accuracy score. We will be able to forecast our project outcomes more precisely thanks to this model, which will help us make better decisions and produce better results.

**Hyperparameter tuning:** We will optimize the hyperparameters of the best-performing models to improve their accuracy.

**Prediction and interpretation:** We will use the final model to make predictions on new data and interpret the results to provide insights for Walmart's business decisions.

Comparing the anticipated sales with the actual sales and figuring out the difference is one method of estimating the possible cost reductions. For instance, if the model predicts that a store would make $100,000 in sales for a certain week, but actual sales come in at $80,000, it is likely that the store was overstocked and the business could have avoided spending $20,000 on wasteful inventory expenditures.

Similar to the last example, it is possible that the store was understocked and the business lost out on $20,000 in potential income if the model forecasts sales of $80,000 but the actual sales end up being $100,000.The cost of the programme for forecasting weekly sales would ultimately rely on a number of variables and presumptions, but if the model is correct and applied properly, there might be considerable cost savings.

When using the model in practise, it is crucial to take these risks into account because there are dangers connected with depending too much on the model's predictions.

**Benefits:**

* Better sales forecasting: Walmart can better manage inventory, optimise pricing schemes, and boost overall sales performance with precise forecasts of weekly sales.
* Cost savings: Walmart may avoid stockouts, eliminate overstocking, and save waste by employing precise sales estimates, which can lower costs and boost profitability.
* Increasing customer happiness: Walmart can guarantee that consumers can locate the things they want when they require them by improving inventory management and more accurately forecasting sales. This will boost customer contentment and loyalty.

**Risks:**

* Overfitting: Using intricate models like random forests and neural networks carries the danger of overfitting, which can result in worse performance on fresh, untried data.

**Timeline:** Our proposed timeline for completing the project is as follows:

Week 1: Data cleaning and preparation

Week 1: Exploratory data analysis and feature engineering

Week 2: Model selection and evaluation

Week 2: Hyperparameter tuning and final model selection

Week 3: Prediction and interpretation

Week 3: Report writing and presentation.

**Deliverables:** The deliverables for this project will include:

A detailed report outlining the methodology, results, and insights gained from the analysis.

A presentation summarizing the key findings and recommendations for Walmart's business decisions.

Python code implementing the data cleaning, exploratory analysis, feature engineering, model selection, hyperparameter tuning, and prediction steps.