DATATHON – SYNCHRONY FINANCIAL

From our analysis and regression, we have obtained an accuracy of around **87.23% (It sometimes fluctuates to 86.xx% while rerunning the same program in the same condition)** by using the **Random_forest classifier**, using different combinations of features with the significant variables mentioned below that yields the best performance:

1. Mortgage Rate
2. New Family Homes for Sale in the United States
3. Total Vehicle Sales
4. Saving and investment-Disaster losses- Private Households and institutions

We also trained and tested the data using other classifiers like SVM Classifier, Baysian Classifier, Decision Tree Regressor and Gradient Boosting Regressor, getting an accuracy between 80-90% for each iteration. Each time we tried eliminating an insignificant variable, the accuracy kept variating.

Throughout the process, one thing we observed was that for all the algorithms, we got the same set of significant variables.

1. **SVM Classifier**- 86% -> Federal Funds, New Family Homes for Sale in the United States, total vehicle sales, Saving and investment-Disaster losses- Private Households and institutions
2. **Bayesian Classifier**- 84.7% -> Real Disposable income, Personal consumption expenditures, Saving and investment-Disaster losses- Private Households and institutions
3. **Decision tree Regressor**- 85.22% -> Mortgage, New Family Homes for Sale in the United States, vehicle sales, Personal consumption expenditures
4. **Gradient Boosting Regressor**- 86.59% -> Mortgage, New Family Homes for Sale in the United States, Real Disposable income, Consumer Price Index

We finally arrived at the highest accuracy of 87.23% above with an RMSE -> 0.12.

Thus, from our analysis, we conclude that the home improvement spending is not significantly affected by an increase in the Fed interest rates for this model. There is only one model (SVM) which gives us Federal funds as a significant variable. Thus, it might probably be positively correlated with home improvement spending, but our optimized model does not classify it as significant enough.

Below is the table representing our column names in the dataframe, and what their meanings are.

Column Names	Meaning	Units	Additional Explanation
ASPNHSUS	Average Sales Price for New Houses Sold in the United States	Dollars	
UNRATE	Civilian Unemployment Rate	Percent	
USACPIHOUMINMEI	CPI (Consumer Price Index)	Index 2010=100	Not Seasonally Adjusted
FEDFUNDS	Effective Federal Funds Rate	Percentage	
CSUSHPINSA	S&P-Case-Shiller U.S. National Home Price Index	Index Jan 2000=100	Not Seasonally Adjusted
MSPNHSUS	Median Sales Price for New Houses Sold in the United States	Dollars	
MSACSR	Monthly_Supply_of_Houses	Months	
MORTGAGE30US	Mortgage	Percentage	
NHFSEPUC	New Houses for Sale by Stage of Construction	Thousands	Under Construction
NHSDPC	New Houses Sold by Stage of Construction_completed	Thousands	Completed
NHSDPNS	New Houses Sold by Stage of Construction_not started	Thousands	Not Started
HNFSEPUSSA	New One Family Homes for Sale in the United States	Thousands	for Sale
HSN1F	New One Family Houses Sold	Thousands	Sold
DSPIC96	Real disposable income	Billion	
TOTALSA	Total Vehicle Sales	Millions	
DATE	Date	“DD/MM/YY”	
DIFSRC1Q027SBEA	Personal consumption expenditures:Financial services and insurance	Billions	Financial services and insurance
W774RC1Q027SBEA	Saving and investment-Disaster losses-Private Households and institutions	Billions	Quarterly
Sales in $MM	Total Sales	Millions	