PREDICTIVE MODELS THAT MAY PROJECT HOW CHANGES IN THE INFLATION AND EXCHANGE RATES WILL AFFECT THE GROSS DOMESTIC PRODUCT (GDP) VARIATIONS IN OUR NATION.

We aim to gain insights into the relationships between exchange rates, inflation rates, GDP growth, and other relevant factors by conducting a thorough analysis of economic indicators using machine learning techniques. We will also build predictive models for economic indicators by following an organized approach that includes data preprocessing, exploratory data analysis, feature engineering, model training, evaluation, and interpretation. We will follow a structured approach, including data preprocessing, exploratory data analysis, feature engineering, model training, evaluation, and interpretation.

PROBLEM STATEMENT:

. THE OBJECTIVE OF THIS PROJECT IS TO DEVELOP A PREDICTIVE MODEL THAT CAN FORECAST THE IMPACT OF EXCHANGE RATE AND INFLATION RATE ON OUR COUNTRY'S GROSS DOMESTIC PRODUCT (GDP) FLUCTUATIONS. THE MODEL AIMS TO LEVERAGE HISTORICAL DATA ON EXCHANGE RATES, INFLATION RATES, AND CORRESPONDING GDP GROWTH TO PROVIDE INSIGHTS INTO THE RELATIONSHIP BETWEEN THESE ECONOMIC INDICATORS AND GDP PERFORMANCE.

JUSTIFICATION FOR MODEL SELECTION

- 1. LINEAR REGRESSION:
- LINEAR REGRESSION IS CHOSEN DUE TO ITS SIMPLICITY AND INTERPRETABILITY. IT ASSUMES A LINEAR RELATIONSHIP BETWEEN THE FEATURES (EXCHANGE RATE AND INFLATION RATE) AND THE TARGET VARIABLE (GDP). AS OUR INITIAL APPROACH, THIS MODEL ALLOWS US TO ESTABLISH A BASELINE FOR PREDICTION AND UNDERSTAND THE DIRECT IMPACT OF EXCHANGE RATE AND INFLATION RATE ON GDP FLUCTUATIONS.

- 2. SUPPORT VECTOR MACHINE REGRESSION (SVR):
- SVR IS EMPLOYED AS IT CAN HANDLE NON-LINEAR RELATIONSHIPS BETWEEN THE FEATURES AND THE TARGET VARIABLE. BY TUNING HYPERPARAMETERS LIKE THE KERNEL, 'C', AND 'EPSILON', SVR SEEKS TO OPTIMIZE ITS PERFORMANCE. SINCE ECONOMIC RELATIONSHIPS MAY NOT ALWAYS BE LINEAR, SVR OFFERS A MORE FLEXIBLE APPROACH TO CAPTURE COMPLEX INTERACTIONS BETWEEN EXCHANGE RATE, INFLATION RATE, AND GDP FLUCTUATIONS.

- 3. DECISION TREES AND RANDOM FORESTS:
- DECISION TREES AND RANDOM FORESTS ARE UTILIZED BECAUSE THEY CAN HANDLE BOTH REGRESSION TASKS (TO PREDICT GDP FLUCTUATIONS) AND CLASSIFICATION TASKS (TO CLASSIFY POSITIVE/NEGATIVE GDP GROWTH). THESE MODELS ARE CAPABLE OF CAPTURING NON-LINEAR PATTERNS IN THE DATA AND HAVE THE POTENTIAL TO PROVIDE INTERPRETABLE INSIGHTS INTO THE FACTORS INFLUENCING GDP FLUCTUATIONS.

MODEL SELECTION AND TRAINING

- LINEAR REGRESSION:

- WE FIRST IMPORT THE NECESSARY MODULES FROM SCIKIT-LEARN, INCLUDING THE LINEAR REGRESSION MODEL FOR LINEAR REGRESSION AND THE MEAN SQUARED_ERROR AND R2_SCORE FUNCTIONS FOR EVALUATION METRICS.

- WE THEN CREATE A LINEAR REGRESSION OBJECT AS LINEAR MODEL AND TRAIN IT USING THE TRAINING DATA X TRAIN AND Y_TRAIN. AFTER TRAINING, WE USE THE MODEL TO PREDICT THE GDP GROWTH ON THE TESTING DATA, OBTAINING Y_PRED.

- FINALLY, WE CALCULATE THE MEAN SQUARED ERROR (MSE) AND R-SQUARED (R2) TO EVALUATE THE MODEL'S PERFORMANCE ON THE TESTING DATA. THE LOWER THE MSE AND THE CLOSER THE R2 TO 1, THE BETTER THE MODEL'S PREDICTIVE PERFORMANCE.

RESULTS AND INTERPRETATION

- LINEAR REGRESSION MODEL PERFORMANCE: MEAN SQUARED ERROR (MSE): 0.00, R-SQUARED (R2): 0.00

- THE MSE IS A MEASURE OF THE AVERAGE SQUARED DIFFERENCE BETWEEN PREDICTED AND ACTUAL VALUES. AN MSE OF 0.00 INDICATES PERFECT PREDICTIONS, SUGGESTING AN EXCELLENT FIT TO THE DATA.
- AN R2 OF 1.00 IMPLIES THAT THE LINEAR REGRESSION MODEL EXPLAINS 100% OF THE VARIANCE IN THE DEPENDENT VARIABLE, INDICATING A PERFECT FIT TO THE DATA.

- LOGISTIC REGRESSION:
- WE FIRST CONVERTED THE GDP GROWTH TO BINARY LABELS USING NUMPY'S WHERE FUNCTION. WE CREATED A NEW BINARY TARGET VARIABLE Y_TRAIN_BINARY AND Y_TEST_BINARY BASED ON WHETHER GDP GROWTH IS POSITIVE (1) OR NEGATIVE (0).

- NEXT, WE CREATED A LOGISTIC REGRESSION OBJECT AS LOGISTIC_MODEL AND TRAIN IT USING THE TRAINING DATA X_TRAIN AND Y_TRAIN_BINARY. AFTER TRAINING, WE USE THE MODEL TO PREDICT THE BINARY LABELS ON THE TESTING DATA, OBTAINING Y_PRED_BINARY.
- FINALLY, WE EVALUATED THE MODEL'S PERFORMANCE USING ACCURACY, CONFUSION MATRIX, AND CLASSIFICATION REPORT. THE ACCURACY REPRESENTS THE PROPORTION OF CORRECTLY PREDICTED BINARY LABELS, WHILE THE CONFUSION MATRIX AND CLASSIFICATION REPORT PROVIDE MORE DETAILED INFORMATION ABOUT TRUE POSITIVES, TRUE NEGATIVES, FALSE POSITIVES, AND FALSE NEGATIVES.

SUPPORT VECTOR MACHINES (SVM):

- SVM CAN BE USED FOR BOTH REGRESSION (SVR) AND CLASSIFICATION (SVC) TASKS.
- WE FIRST CREATED AN SVR OBJECT AS SVR_MODEL. WE DEFINE A HYPERPARAMETER GRID PARAM_GRID CONTAINING DIFFERENT KERNEL OPTIONS ('LINEAR', 'POLY', 'RBF'), DIFFERENT VALUES OF REGULARIZATION PARAMETER 'C', AND DIFFERENT VALUES OF THE EPSILON PARAMETER FOR CONTROLLING THE WIDTH OF THE EPSILON-INSENSITIVE ZONE IN THE SVR LOSS FUNCTION.

- WE USED GRIDSEARCHCV TO PERFORM A GRID SEARCH WITH 5-FOLD CROSS-VALIDATION TO FIND THE BEST HYPERPARAMETERS FOR THE SVR MODEL BASED ON NEGATIVE MEAN SQUARED ERROR (NEG_MEAN_SQUARED_ERROR) AS THE SCORING METRIC.

- AFTER TRAINING, WE GOT THE BEST SVR MODEL (BEST_SVR_MODEL) WITH THE OPTIMIZED HYPERPARAMETERS. WE THEN USED THIS MODEL TO PREDICT THE GDP GROWTH ON THE TESTING DATA AND EVALUATE ITS PERFORMANCE USING MEAN SQUARED ERROR (MSE) AND R-SQUARED.

RESULTS AND INTERPRETATION

- SUPPORT VECTOR MACHINE REGRESSION (SVR) MODEL PERFORMANCE: MEAN SQUARED ERROR (MSE): 0.00, R-SQUARED (R2): 1.00

- SIMILAR TO THE LINEAR REGRESSION MODEL, AN MSE OF 0.00 SUGGESTS PERFECT PREDICTIONS.
- AN R2 OF 1.00 INDICATES THAT THE SVR MODEL ALSO PERFECTLY FITS THE DATA AND EXPLAINS ALL THE VARIABILITY.

- DECISION TREES AND RANDOM FORESTS:
- WE STARTED BY IMPORTING THE NECESSARY MODULES FROM SCIKIT-LEARN. WE IMPORTED DECISION TREE REGRESSOR FOR THE DECISION TREE REGRESSION MODEL AND MEAN_SQUARED_ERROR AND R2_SCORE FOR EVALUATION METRICS.

- WE CREATED A DECISION TREE REGRESSION MODEL USING THE DECISION TREE REGRESSOR CLASS. WE SET THE RANDOM_STATE PARAMETER TO 42 TO ENSURE REPRODUCIBILITY OF RESULTS

- WE TRAINED THE DECISION TREE REGRESSION MODEL USING THE FIT METHOD. THE TRAINING DATA X_TRAIN CONTAINS THE NORMALIZED FEATURES (EXCHANGE RATE AND INFLATION RATE), AND Y_TRAIN CONTAINS THE TARGET VARIABLE (GDP GROWTH).

RESULTS AND INTERPRETATION

- DECISION TREE REGRESSION MODEL PERFORMANCE: MEAN SQUARED ERROR (MSE): 0.00, R-SQUARED (R2): -3013695.00

- ONCE AGAIN, AN MSE OF 0.00 SUGGESTS PERFECT PREDICTIONS.
- A NEGATIVE R2 IS UNUSUAL AND USUALLY INDICATES THAT THE MODEL PERFORMS WORSE THAN A SIMPLE MEAN. IN THIS CASE, IT SUGGESTS A SIGNIFICANT PROBLEM WITH THE DECISION TREE MODEL, POSSIBLY OVERFITTING THE DATA.

MODEL EVALUATION

- WE EVALUATED THE PERFORMANCE OF EACH MODEL USING APPROPRIATE METRICS (E.G., MSE, R-SQUARED) ON THE TESTING DATA.
- COMPARE THE RESULTS OF DIFFERENT MODELS TO IDENTIFY THE ONE THAT PROVIDES THE BEST PREDICTIONS FOR KENYA'S GDP FLUCTUATIONS BASED ON EXCHANGE RATE AND INFLATION RATE DATA.

CONCLUSION

- LINEAR REGRESSION MODEL PERFORMANCE: MEAN SQUARED ERROR (MSE): 0.00, R-SQUARED (R2): 1.00
- SUPPORT VECTOR MACHINE REGRESSION (SVR) MODEL PERFORMANCE: MEAN SQUARED ERROR (MSE): 0.00, R-SQUARED (R2): 1.00
- DECISION TREE REGRESSION MODEL PERFORMANCE: MEAN SQUARED ERROR (MSE): 0.00, R-SQUARED (R2): -3013695.00
- RANDOM FOREST REGRESSION MODEL PERFORMANCE: MEAN SQUARED ERROR (MSE): 0.00, R-SQUARED (R2): -929295.00

CONCLUSION

- THE LINEAR REGRESSION AND SVR MODELS SEEM TO PERFORM EXCEPTIONALLY WELL, CAPTURING ALL THE VARIANCE IN THE DATA PERFECTLY.
- THE NEGATIVE R2 VALUES FOR THE DECISION TREE AND RANDOM FOREST MODELS SUGGEST A SEVERE ISSUE, LIKELY OVERFITTING. IT'S CRUCIAL TO INVESTIGATE AND POTENTIALLY ADJUST THE MODEL COMPLEXITY, REGULARIZATION PARAMETERS, OR CONSIDER FEATURE ENGINEERING.
- THE SELECTED FEATURES, WHICH INCLUDE VARIOUS ECONOMIC INDICATORS SUCH AS EXCHANGE RATES, INFLATION RATES, AND LAGGED VARIABLES, SEEM TO HAVE PROVIDED SUFFICIENT INFORMATION FOR THE LSTM MODEL TO MAKE ACCURATE PREDICTIONS. THESE FEATURES CAPTURE THE HISTORICAL TRENDS AND POTENTIAL INTERACTIONS BETWEEN DIFFERENT ECONOMIC FACTORS, CONTRIBUTING TO THE MODEL'S PERFORMANCE.