palAI & mAIs

Manejo, Pajarilla, Vito

COMPARATIVE ANALYSIS OF MACHINE LEARNING ALGORITHMS FOR FORECASTING PALAY AND CORN PRODUCTION IN REGION VI (WESTERN VISAYAS)

Manejo, Pajarilla, Vito

CMSC 197 Machine Learning

OUTLINE

Introduction

- Motivations
- Objectives

Methodology

- Dataset and preprocessing
- ML Algorithms and Hyperparameters

Results

Conclusion and Recommendation

Manejo, Pajarilla, Vito

CMSC 197 Machine Learning

MOTIVATION

Growing for a Greener Future

Rice and corn are the two most important staple crops in the Philippines, essential not only for food security but also as a source of livelihood for millions of Filipinos.

Manejo, Pajarilla, Vito

CMSC 197 Machine Learning

MOTIVATION

Growing for a Greener Future

Crop prediction of rice and corn in the Philippines can help in improved food security, enhanced farm management, economic stability, and informed policy formulation.

Manejo, Pajarilla, Vito

CMSC 197 Machine Learning

OBJECTIVE

Growing for a Greener Future

Aims to asses different machine learning algorithms in forecasting time series of Rice and Corn production in the provinces of Western Visayas, Philippines and compare the viability and accuracy of each forecasting.

Manejo, Pajarilla, Vito

METHODOLOGY

Dataset

Preprocessing and Hyperparameter Tuning

ML Models Training

Model Testing

Results

Manejo, Pajarilla, Vito

DATASET

	Croptype	Province	Quarter	Year	Area	Production
840	Irrigated P	Aklan	1	1994	9980	43151
841	Irrigated P	Antique	1	1994	11540	26667
842	Irrigated P	Capiz	1	1994	7300	20440
843	Irrigated P	Guimaras	1	1994	1930	7123
844	Irrigated P	Iloilo	1	1994	30770	118772
845	Rainfed Pa	Aklan	1	1994	8240	14339
846	Rainfed Pa	Antique	1	1994	3010	4635
847	Rainfed Pa	Capiz	1	1994	33890	74558
848	Rainfed Pa	Guimaras	1	1994	3910	8459
849	Rainfed Pa	Iloilo	1	1994	52990	137774
850	Palay	Aklan	1	1994	18220	57490
851	Palay	Antique	1	1994	14550	31302
852	Palay	Capiz	1	1994	41190	94998
853	Palay	Guimaras	1	1994	5840	15582
854	Palay	Iloilo	1	1994	83760	256546
855	White Cor	Aklan	1	1994	2581	6370
856	White Cor	Antique	1	1994	110	46
857	White Cor	Capiz	1	1994	160	91
858	White Cor	Guimaras	1	1994	60	60
859	White Cor	Iloilo	1	1994	1240	1240
860	Yellow Co	Aklan	1	1994	2581	6370

The dataset was sourced from the Philippine Statistics Authority openSTAT website. It encompasses the quarterly production of rice and corn across five provinces in Western Visayas: Iloilo, Aklan, Capiz, Guimaras, and Antique from year 1984 to 2024. The dataset includes information on the area harvested, year, quarter, and crop type.

Manejo, Pajarilla, Vito

PREPROCESSING

Data transformation for time-series suitable format

Removal of outliers and missing data

One hot encoding of categorical variables and log transformation of Production variable

TRAINING DATA

Quarter 1 of 1984 to quarter 4 of 2019

TESTING DATA

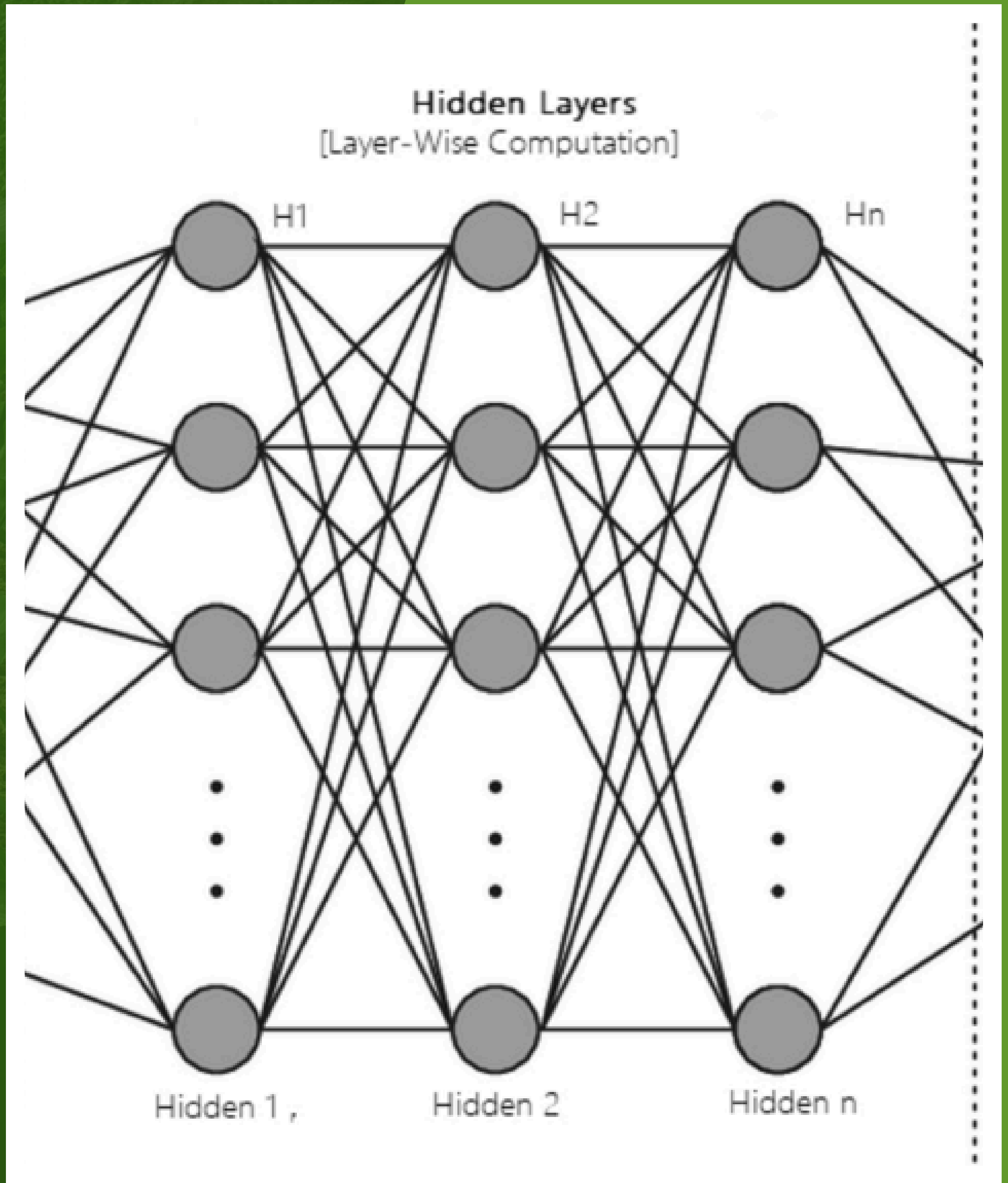
Quarter 1 of 2020 to quarter 2 of 2024

Manejo, Pajarilla, Vito

CMSC 197 Machine Learning

MACHINE LEARNING MODELS

- Linear Regression
- Extreme Gradient Boosting (XGBoost)
- K-Nearest Neighbor (KNN)
- Support Vector Regression (SVR)
- Random Forest (RF) Model
- Artificial Neural Networks (ANN)



Manejo, Pajarilla, Vito

HYPERPARAMETERS

1 XGBOOST	2 RANDOM FORESTS
• *learning rate = 0.2* • *max_depth = 5* • *n_estimators = 500* • *reg_alpha = 0.2* • *reg_lambda = 0.1* • *subsample = 0.8*	• *bootstrap = True* • *max_depth = None* • *max_features = 'log2'* • *min_samples_leaf = 1* • *min_samples_split = 2* • *n_estimators = 200*

Manejo, Pajarilla, Vito

CMSC 197 Machine Learning

HYPERPARAMETERS

3 LINEAR REGRESSION • *No hyperparameters for tuning*	**4 ARTIFICIAL NEURAL NETWORKS** • *None applied*
5 SUPPORT VECTOR REGRESSION • *bootstrap = True* • *max_depth = None* • *max_features = 'log2'* • *min_samples_leaf = 1* • *min_samples_split = 2* • *n_estimators = 200*	**6 K-NEAREST NEIGHBOR** • *metric = 'manhattan'* • *n_neighbors = 3* • *weights = 'distance'*

Manejo, Pajarilla, Vito

METRICS

- Mean Absolute Error (RMSE)
- Root Mean Squared Error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- Symmetric mean absolute percentage error (SMAPE)

MAPE Value.	SMAPE Value	Predictive Performance Evaluation
<10%	<10%	Highly accurate forecasting
10–20%	10–20%	Good forecasting
20–50%	20–50%	Reasonable forecasting
>50%	>50%	Inaccurate forecasting

RESULTS

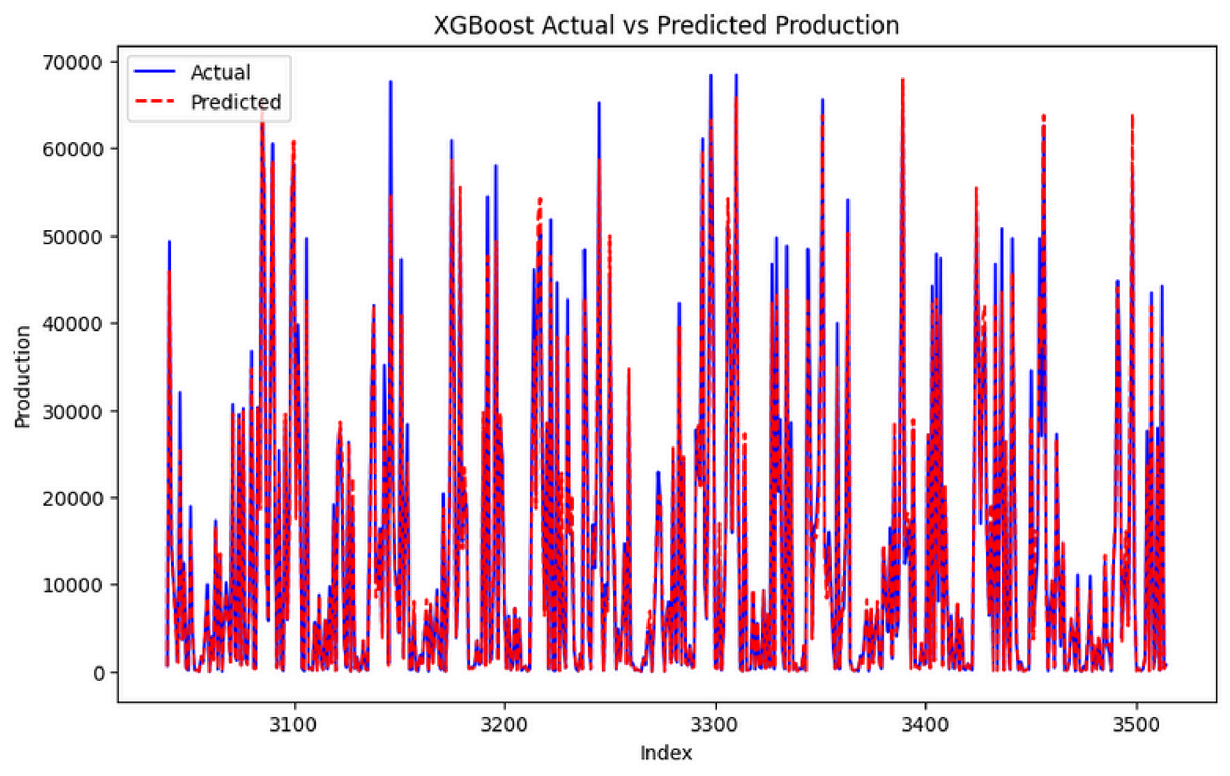
Model	MAE	MAPE	RMSE	SMAPE
XG Boost	1217.253380	14.195011	2225.081848	7.162455
Random Forest	1952.584102	25.146939	3985.710568	10.447967
Linear Regression	16062.700263	133.043564	54028.439415	36.773868
Support Vector Regression	2171.360387	62.052071	3666.856666	19.274818
K-Nearest Neighbor	2276.290483	33.826273	4271.281769	16.806541
ANN	7486.512184	68.621257	17026.412299	19.274818

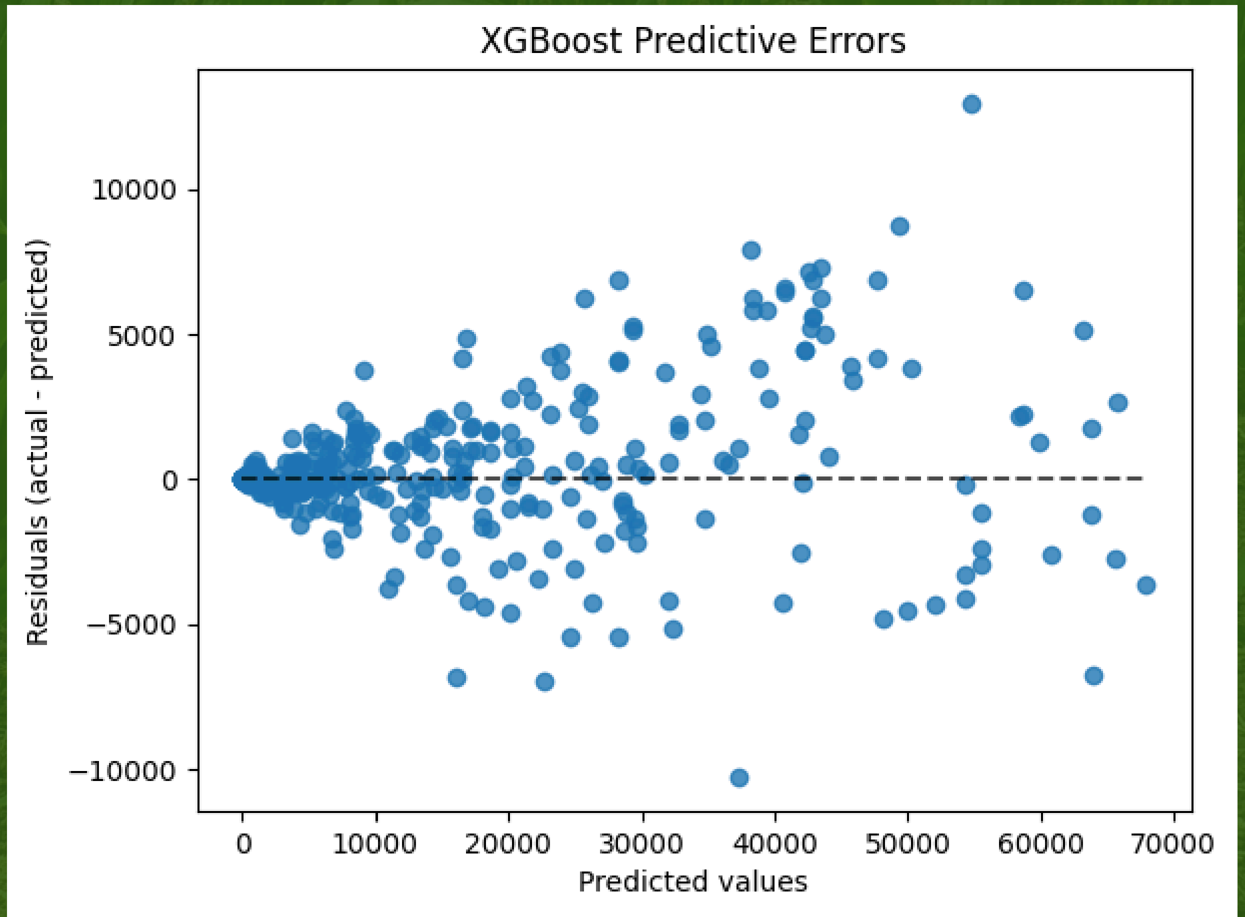
Manejo, Pajarilla, Vito

CMSC 197 Machine Learning

RESULTS

Extreme Gradient Boosting (XGBoost)

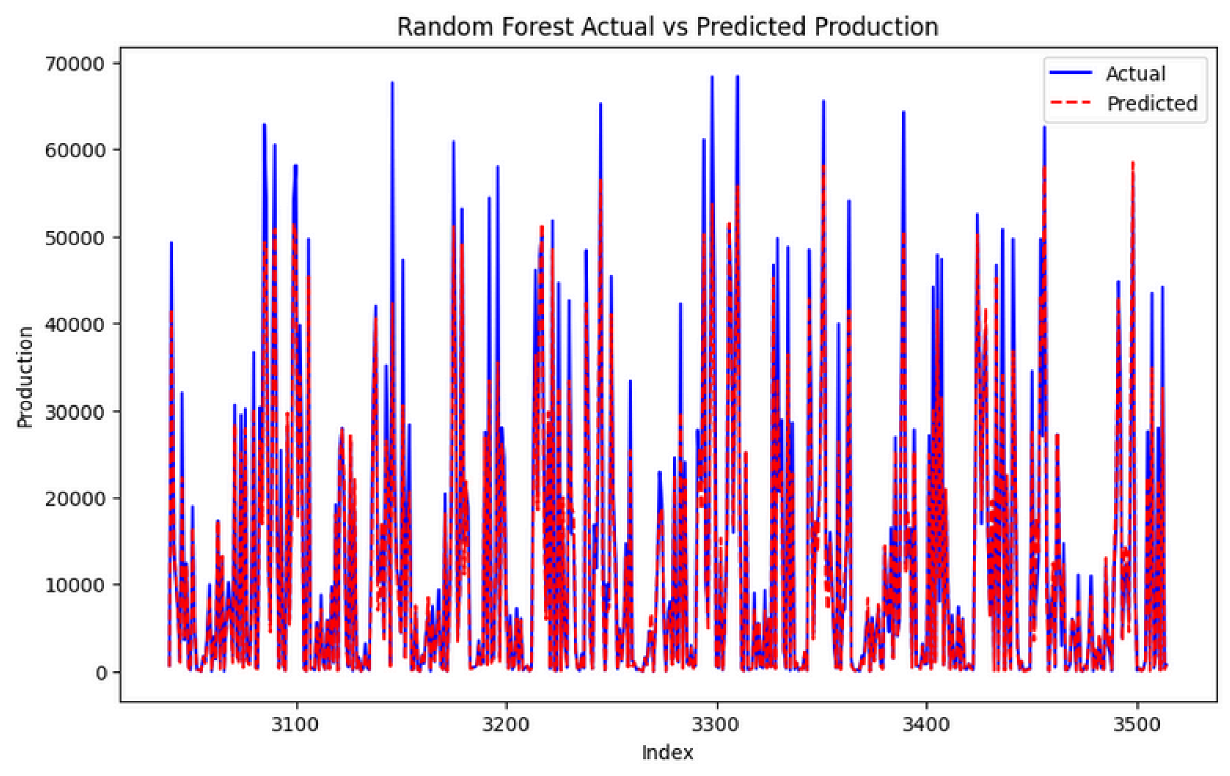


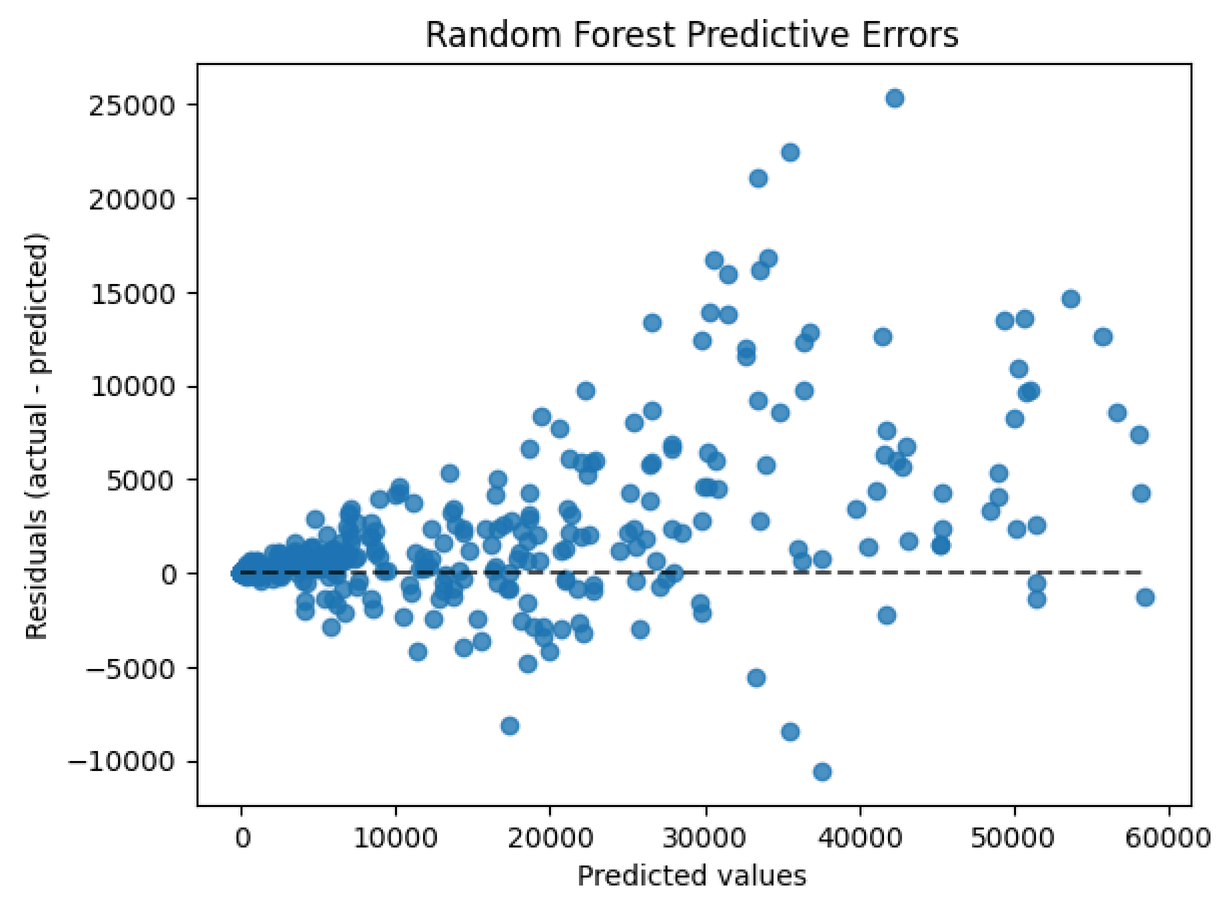


Manejo, Pajarilla, Vito

RESULTS

Random Forest (RF) Model



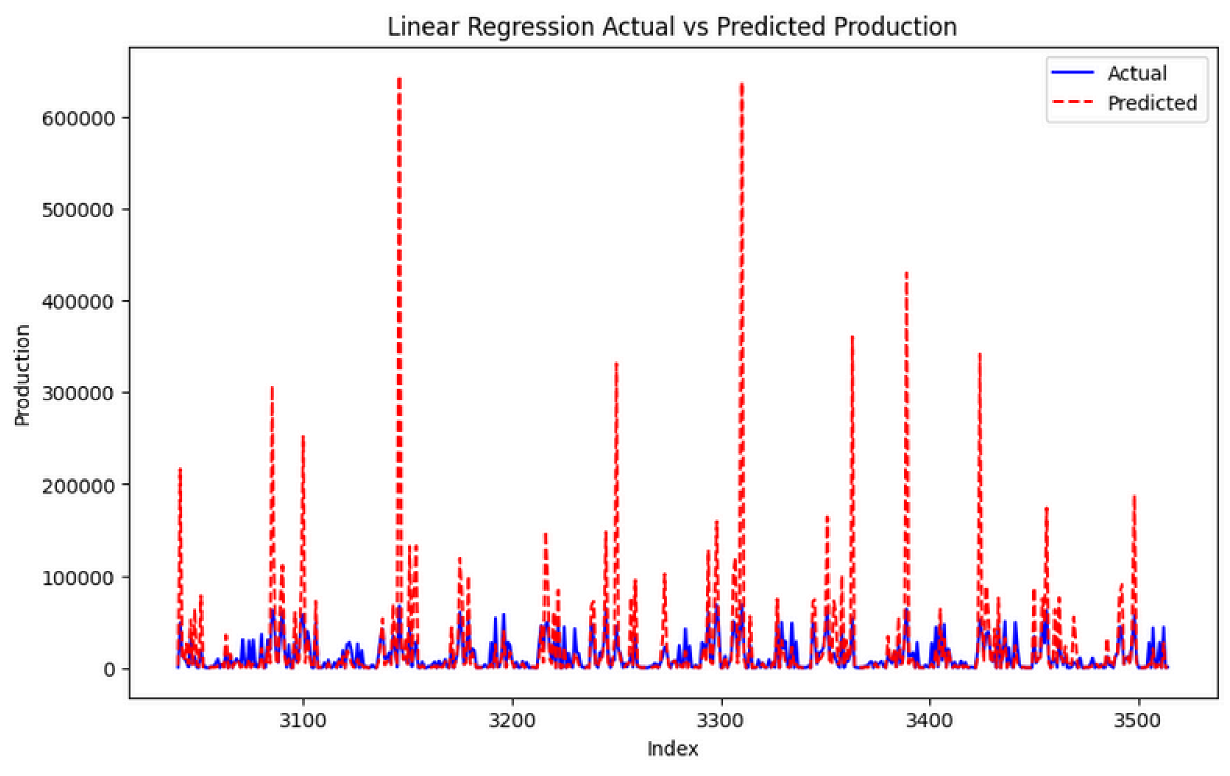


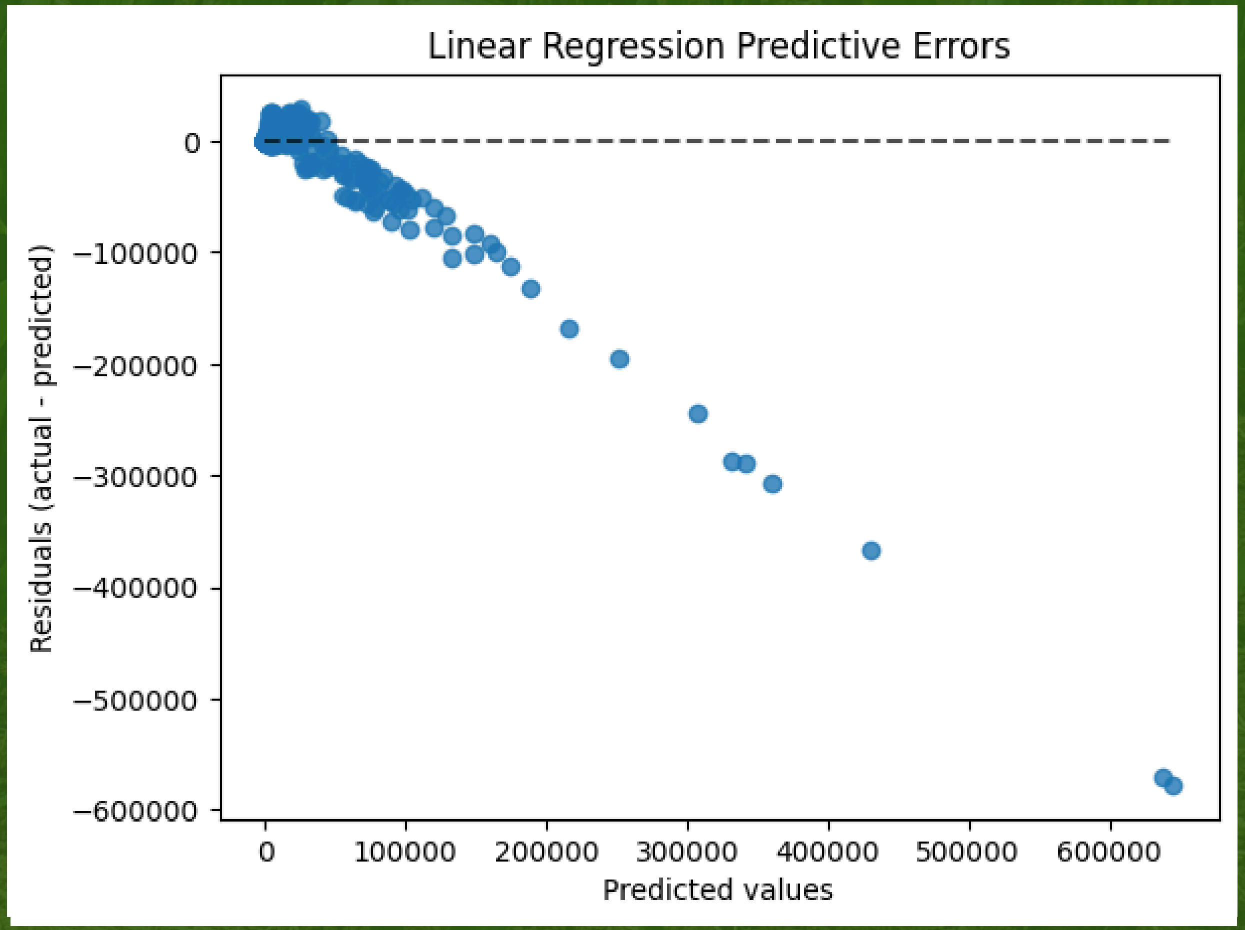
Manejo, Pajarilla, Vito

CMSC 197 Machine Learning

RESULTS

Linear Regression (LR) Model



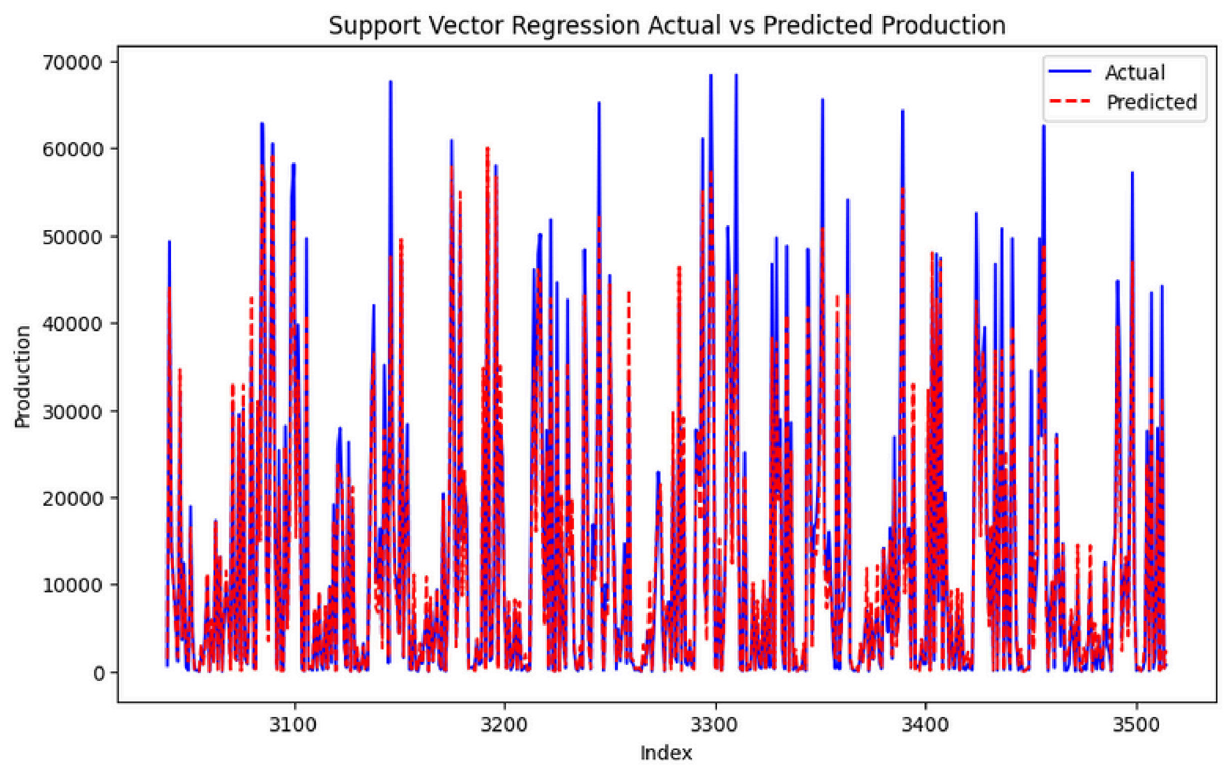


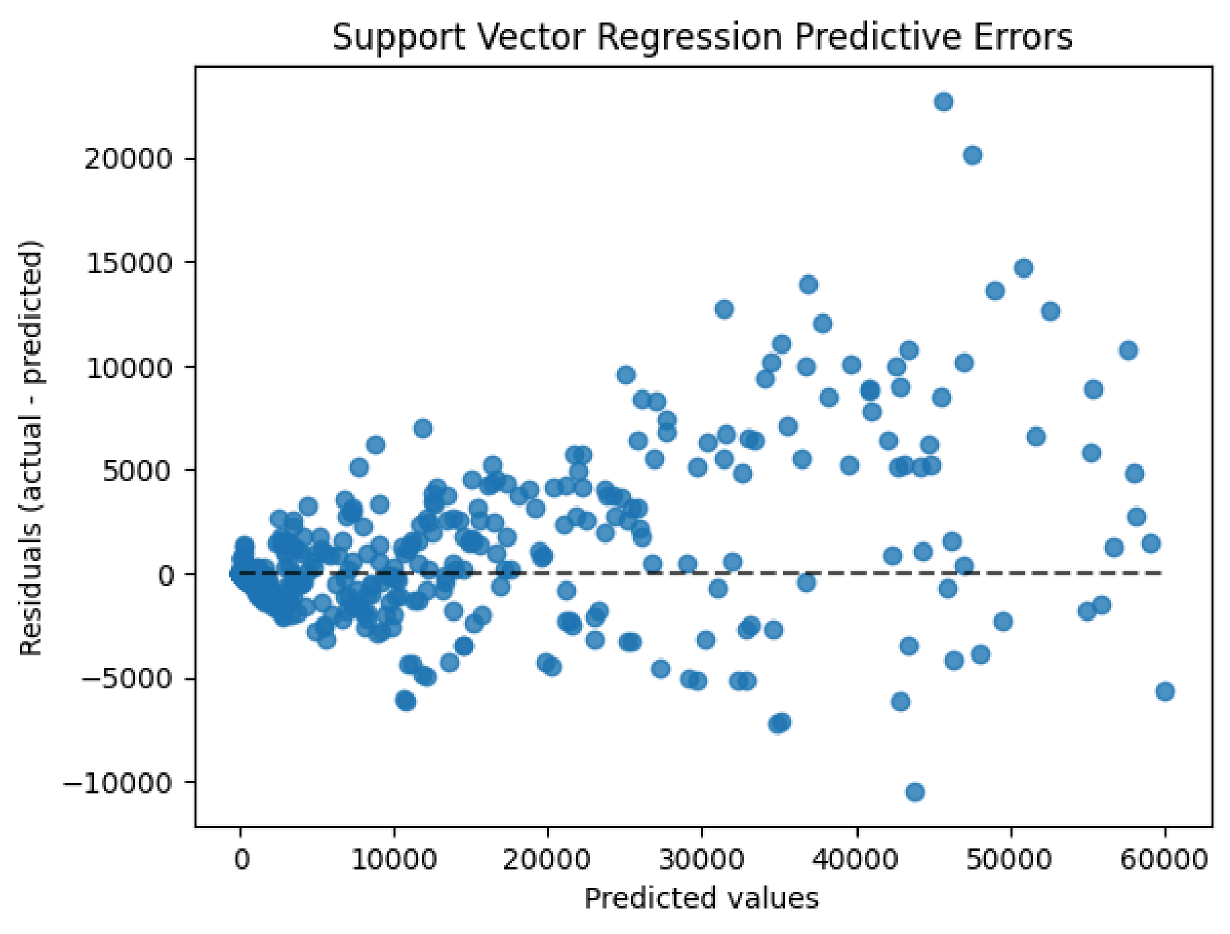
Manejo, Pajarilla, Vito

CMSC 197 Machine Learning

RESULTS

Support Vector Regression (SVR)

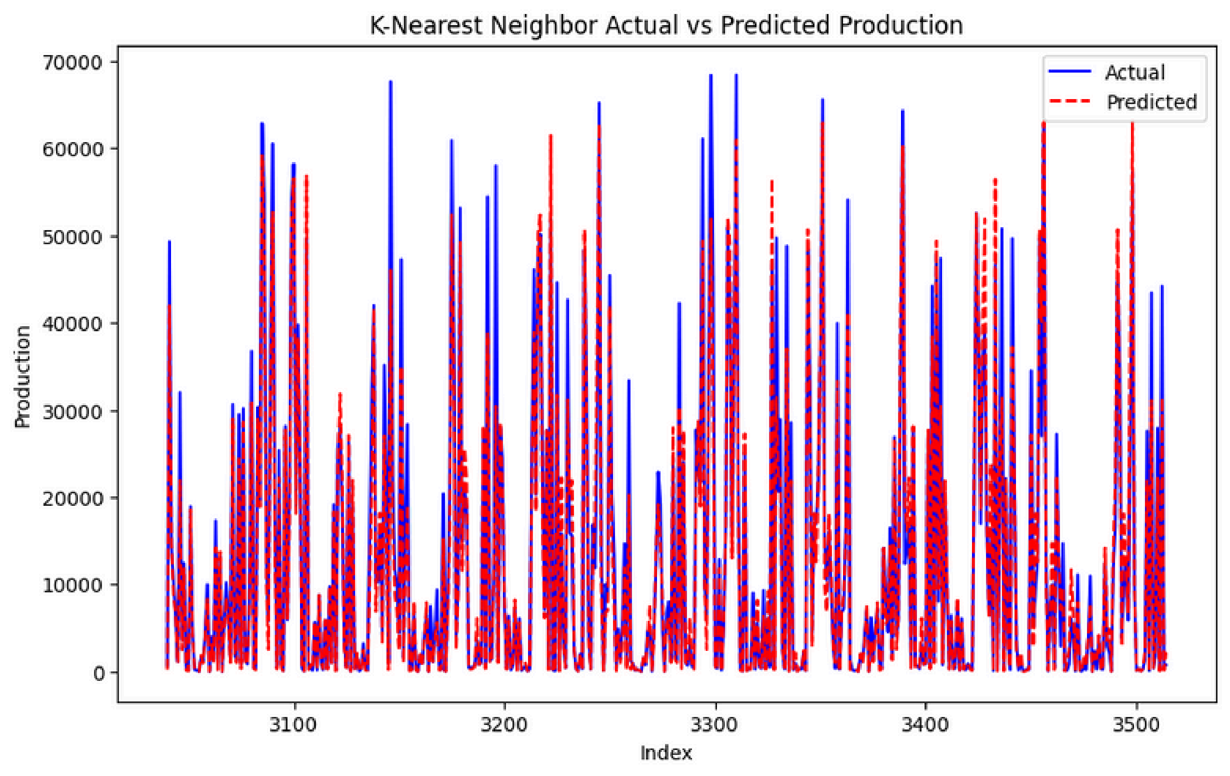


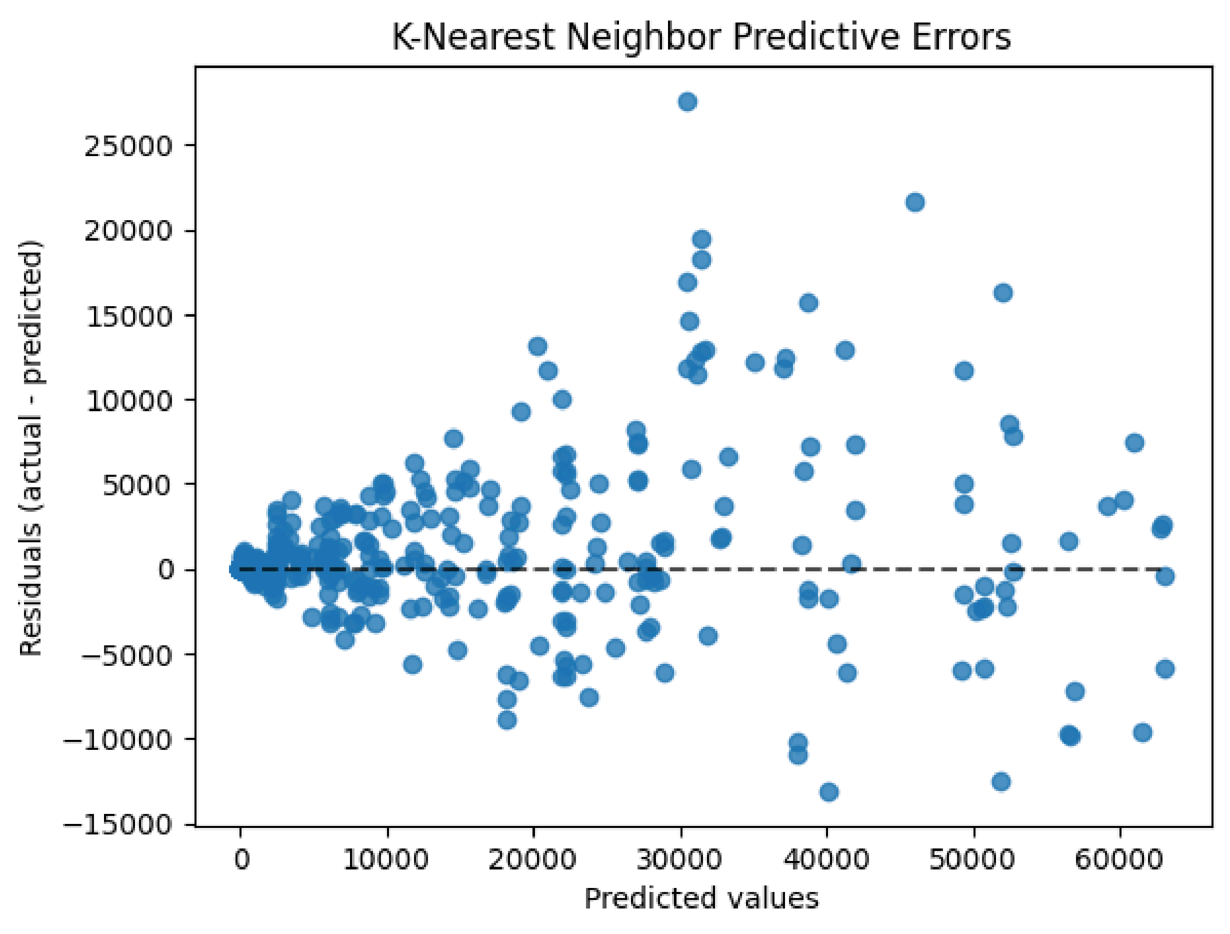


Manejo, Pajarilla, Vito

RESULTS

K-Nearest Neighbor (KNN)

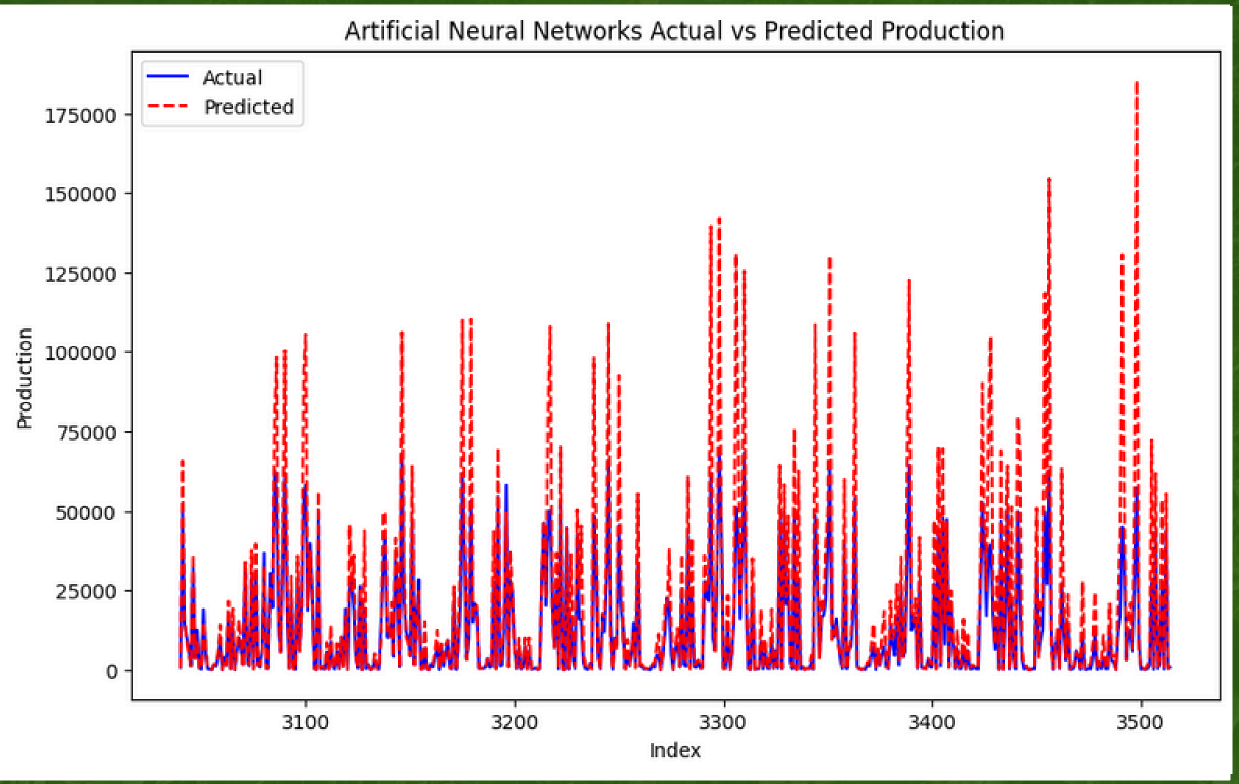


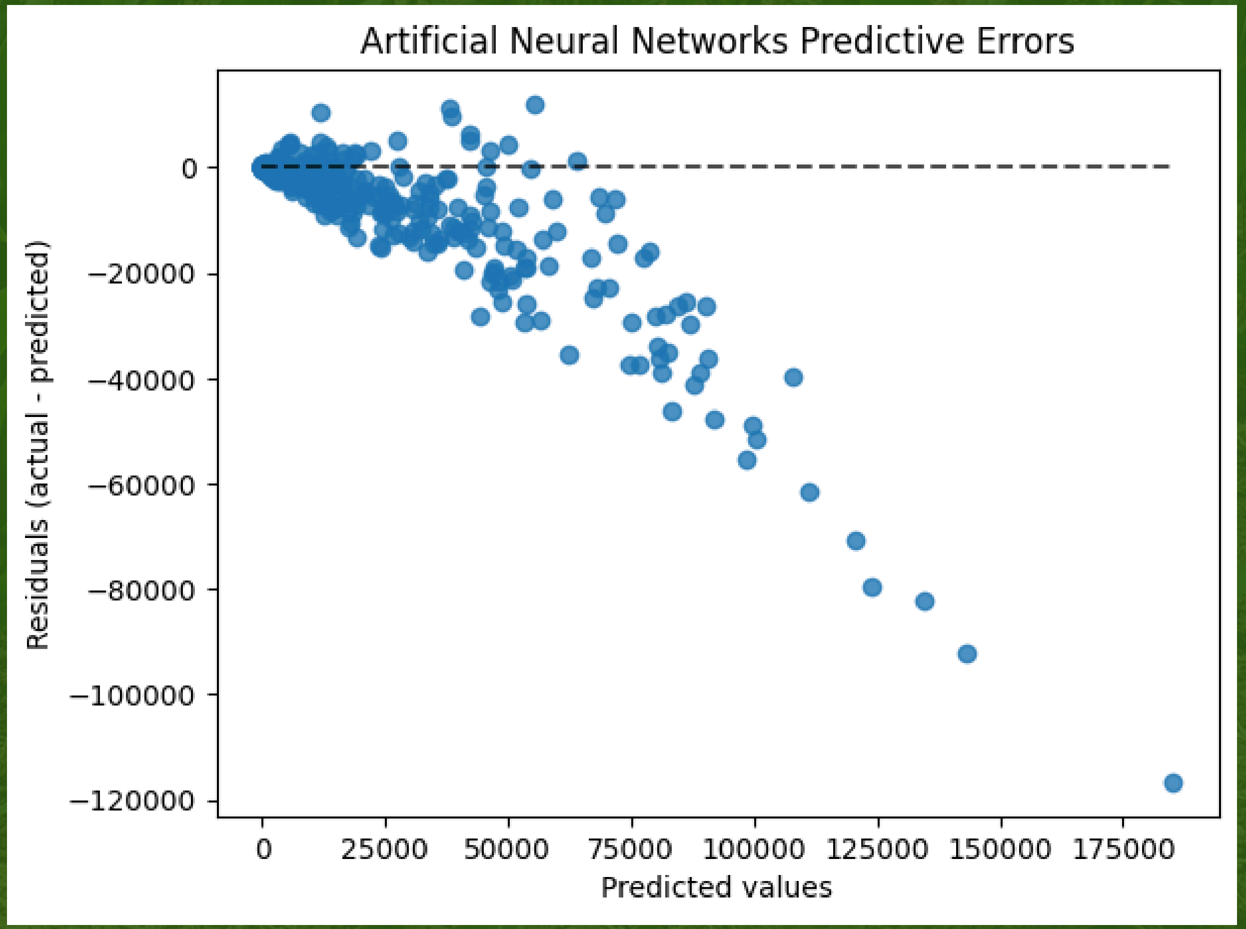


Manejo, Pajarilla, Vito

RESULTS

Artificial Neural Networks (ANN)





Manejo, Pajarilla, Vito

CONCLUSION

1. The XGBoost model was the most accurate in predicting palay and corn production, while linear regression showed the lowest metric results.

2. Developing predictive models using XGBoost is significant for forecasting crop production in the Philippines and is beneficial for future studies in the same context.

Manejo, Pajarilla, Vito

RECOMMENDATIONS

1. Include additional variables, such as soil type, weather conditions, irrigation methods, and agricultural practices, to improve prediction accuracy.

2. Study how the accuracy of the model's predictions and analysis can vary depending on the data and variable settings, and determine which approach is superior in all cases.

3. Investigate time series forecasting algorithms such as Autoregressive Integrated Moving Average (ARIMA) and Seasonal Trend Decomposition to enhance the analysis.

Manejo, Pajarilla, Vito

THANK YOU

CMSC197 Machine Learning

See You Next