MACHINE LEARNING APPROACHES FOR PRODUCTIVITY ESTIMATION OF BULLDOZERS

CASE STUDY FINAL PRESENTATION

TUGBERK ERDOGMUS
SINEM KARAOMEROGLU
SARA MORADI JAFARBEGLOO
UGUR EYLIK

Introduction

01

- **Objective:** Explore various machine learning approaches for estimating bulldozer productivity
- **Data Generation Techniques:** Uniform distribution, Monte Carlo simulation, Latin hypercube sampling
- **Machine Learning Models:** Linear regression, random forest, ridge regression, lasso regression, Elastic Net regression

Phase 1 - Methodology & Results

02

- **Dataset Size:** 1000 data points
- **Techniques Used:** Uniform distribution, Monte Carlo simulation, Latin hypercube sampling
- **Models Evaluated:** Linear regression, random forest

Model	R^2	MSE
Uniform Distribution with Linear Regression	0.9160	4.1119
Uniform Distribution with Random Forest	0.8055	9.5275
Monte Carlo Simulation with Linear Regression	0.9160	4.1119
Monte Carlo Simulation with Random Forest	0.8055	9.5275
Latin Hypercube Sampling with Linear Regression	0.9234	5.3756
Latin Hypercube Sampling with Random Forest	0.8195	11.2527

Phase 2 - Methodology & Results

03

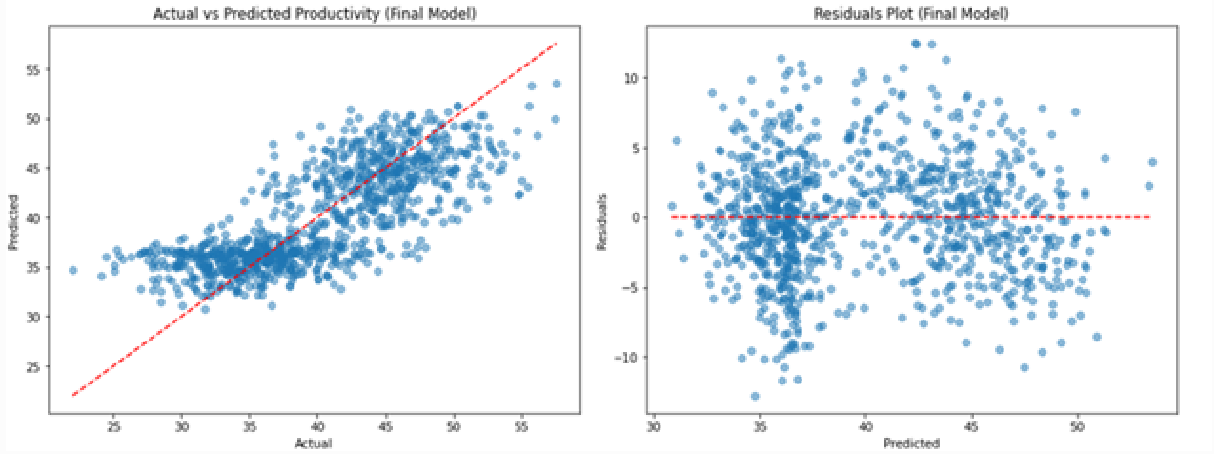
- **Dataset Size:** Increased to 5000 data points
- **Additional Techniques:** Feature engineering and preprocessing (encoding types, VIF analysis, polynomial features, hyperparameter tuning with CV)
- **Models Evaluated:** Linear regression, random forest, ridge regression, lasso regression, Elastic Net regression

Model	R^2	MSE
Uniform Distribution with Linear Regression	0.6910	20.4564
Uniform Distribution with Random Forest	0.7486	16.6451
Uniform Distribution with Ridge Regression	0.6912	20.4442
Uniform Distribution with Lasso Regression	0.6914	20.4325
Uniform Distribution with Elastic Net Regression	0.6914	20.4339
Monte Carlo Simulation with Linear Regression	0.5900	17.4087
Monte Carlo Simulation with Random Forest	0.7486	16.6451
Monte Carlo Simulation with Ridge Regression	0.5900	17.4116
Monte Carlo Simulation with Lasso Regression	0.5891	17.4482
Monte Carlo Simulation with Elastic Net Regression	0.5891	17.4487
Latin Hypercube Sampling with Linear Regression	0.7058	18.8080
Latin Hypercube Sampling with Random Forest	0.7239	16.5157
Latin Hypercube Sampling with Ridge Regression	0.6991	18.7784
Latin Hypercube Sampling with Lasso Regression	0.7168	18.9227
Latin Hypercube Sampling with Elastic Net Regression	0.6976	18.8724

Strengths & Limitations

04

- **Strengths:**
  - Comprehensive comparison of data generation techniques and machine learning models
  - Enhanced performance with feature engineering and preprocessing techniques
- **Limitations:**
  - Synthetic nature of data may not capture real-world nuances



Actual vs Predicted Productivity and Residuals Plot for Linear Regression Model on Monte Carlo Simulation Data

Conclusion

05

- Better results with Linear Regression when the dataset is less complex and easier to implement and interpret
- The Linear Regression with Latin Hypercube Sampling yielded the best overall performance when data was 1000
- Random Forest is more flexible and better at modeling complex patterns in the data
- The Random Forest with Uniform Distribution yielded the best overall performance when data was 5000
- As the amount of our data increases, Monte Carlo Simulation creates a more dispersed dataset and makes it more difficult to build a model on it

Thank You