Assignment 2: Build Regression Models

Anuj Bansal

Data Sciences, National University

DDS-8555 v1: Predictive Analysis

Dr. Mohammad

July 25th, 2025

This week’s report goes over building and comparing regression models (Regression in machine learning, 2025).  We studied data analysis and how to engineer data so that it is suited for model building. We looked at the output of the model to determine which model fits better with data.

When answering questions from the ISLR python (Gareth James, 2023) chapter 3, we observed the importance of coefficients to write regression equation. Further, we learnt about p-value to understand how to look for significant variables. we used R squared and Adjusted R squared values to understand the coverage of model to the data.

The next part of the learning cleansed data and built 2 regression models (linear regression and random forest (Geek for Geek, 2025)). MSE and RMSE values were used to select a better model. Lastly data was generated with the selected model for submission. Lastly, all was submitted to Kaggle (appendix A).

Appendix A

A screenshot of a computer

AI-generated content may be incorrect.

# References

*Classification Model* . (2023, sep 30). Retrieved from medium: https://medium.com/@karan.kamat1406/which-classification-model-should-you-use-a-cheat-sheet-for-machine-learning-practitioners-3fea0bcab04e

Gareth James, D. W. (2023). *An Introduction to Statistical Learning: with Applications in Python.* Springer. Retrieved from https://www.statlearning.com/

*Regression in machine learning*. (2025, Jan 13). Retrieved from geek for geeks: https://www.geeksforgeeks.org/machine-learning/regression-in-machine-learning/

RITHP. (2023, jan 4). *Medium*. Retrieved from Improving Precision and Recall in Machine Learning: Tips and Techniques: https://medium.com/@rithpansanga/improving-precision-and-recall-in-machine-learning-tips-and-techniques-acb5a5fd27a6

scikit-learn developers . (2007-2025). *Metrics and scoring: quantifying the quality of predictions*. Retrieved from scikit-learn: https://scikit-learn.org/stable/modules/model\_evaluation.html#classification-metrics