

**THE APPLICATION OF MACHINE LEARNING TO
PREDICT THE REVENUE & SIGNIFICANT FACTORS
THAT IMPACT IT WITH SPECIAL REFERENCE TO
DESSFORD ESTATE**

Lasanthi Ladduwahetty

A dissertation submitted in partial fulfilment of the requirement for
Master of Science degree in Business Analytics

**Department of Business
Informatics Institute of Technology, Sri Lanka
in collaboration with
Robert Gordon University, UK**

2022

Abstract

Machine learning techniques are becoming increasingly popular around the world. However, when considering Sri Lanka's Tea industry; this area or its application is very limited. The aim of this study is to use machine learning algorithms to predict the Revenue at Dessford Estate. The objectives of this research are to Identify the main factors that affect revenue and identify how it has impacted over time, building a model to Predict the revenue of the Business, identify the most suitable model to predict the revenue at Dessford Estate and its relationship with the other variables and lastly to identify the significant cost contributions and provide a roadmap or recommendations to Dessford through the use of analytics.

The data was provided from year 2011 till 2021 from Dessford Estate with formal consent to be utilized for this thesis. 7 Machine Learning Regression models were implemented for conduct the analysis - Multiple Linear Regression, Ridge Regression, LASSO Regression, Elastic Net Regression, Decision Tree Regression, XG Boost and K-Nearest Neighbor. Two Strategies were followed to analyze the results of these respective Regression models. Out of the models implemented the best 3 performing algorithms were Ridge Regression, LASSO Regression, Elastic Net Regression.

Furthermore, Four Time Series Forecasting models were implemented to analyze and predict Revenue based on past trends, seasonality of the data. Simple Exponential Smoothing (SES), Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Moving Average (SARIMA) and Holt Winters were analyzed. Out of which SARIMA was selected as the best performing model due to having the least number of error values.