PROJECT SYNOPSIS

Name of the student	:	Oswald Aguiar
Address for Correspondence	:	A – 701 Star Classic Chs, Vasai Road (West), Mumbai, India
Email	:	aguiaroswald73@gmail.com
Mobile	:	+971 544566757

Registration No.	:	P392010001
Program	:	Post Graduate Certificate in Management (Business Analytics)
Institute	:	Center for Management Education, All India Management Association, New Delhi
Working Title	:	Predicting Water Pump Failure
Project Guide	:	Mr. Saurabh Agarwal B.Tech(IIT Kanpur), M.Tech(IIT Delhi) Cell:9415077960, WhatsApp: 9519166622
Designation & Address	:	Faculty & Expert, HBTI, E&ICT Academy, Great Learning, NMIMS, AIMA Kanpur, UP, India

SYNOPSIS CONTENTS

1. WORKING TITLE

Predicting Water Pump Failure

2. PURPOSE OF THE RESEARCH

Water is fundamental to life and the environment; it plays a central role in both, economic and social development activities. Water touches all the spheres of human life including: domestic, livestock, fisheries, wildlife, industry and energy, recreation and other social — economic activities. It plays a pivotal role in poverty alleviation through the enhancement of food security, domestic hygiene and the environment. The availability of safe and clean water raises the standard of living while the inadequacy of it poses serious health risks and leads to the decline in the living standards and life expectancy.

Major fresh water sources in Tanzania include: lakes, rivers, streams, dams and groundwater. However, these are not well distributed all over the country. Some areas lack both surface and ground water sources.

Increasing population growth and urbanization pose serious pressure on the quantity and quality of available water. The sustainability of the present and future human life and environment depends mainly on proper water resources management.

This starts with analyzing the functional status of the available water points in Tanzania. In this analysis I will be looking into the dataset of water pumps in Tanzania to predict the operating condition of a water point.

By predicting the pumps which are functional but needs repair, decreases the overall cost for the Tanzanian Ministry of Water. Which can improve the maintenance operations of the water pumps and make sure that clean, potable water is available to communities across Tanzania.

3. DATA FOR THE PROJECT

The data for this project was sourced from the DrivenData.org the dataset is accessible via the following web link:

https://www.drivendata.org/competitions/7/pump-it-up-data-mining-the-water-table/page/25/

The data set is comprised of attributes that describe a total of 59,400 Tanzanian water pumps. Each water pump is represented by a total of 39 qualititative and quantitative attributes that describe such things as the type of pump, location, altitude, installation funding source, management method, year constructed, water source, and the quality of the water delivered by the pump.

4. AIM

To predict one of three possible values:

- `functional`: The pump is operational and not in need of maintenance
- `functional needs repair`: The pump is operational but is in need of repair
- `non functional`: The pump is inoperable

5. OBJECTIVE

Following are the objectives of the project

- To discover factors that are responsible for pump performance
- Under what circumstances are the pumps likely to fail
- Comparison of model outputs and explainability
- What can be done to improve pump performance

6. KEY QUESTIONS

i. What are the primary predictors of pump failure?
ii. Is there a parity in the performance of water pumps funder by the government and other?
iii. Is there a trend in performance of pumps over time?
iv. Can a model be developed to ensure pumps remain functioning longer?

7. METHODOLOGY

The components of the Methodology are enumerated below.

- Collect the data from relevant sources
- Creating training and Testing dataset
- Data Cleaning:
  - Treating missing values
  - Outlier treatment
  - Data transformation
- Univariate Analysis
- Multivariate Analysis
- Feature Engineering
- Building a Predictive Model
- Evaluating Model Performance
- Interpretation of Model Output

8. ALGORITHMS

It is decided to explore the use of the following algorithms for the development of the model

1. Linear Discriminant Analysis (LDA)
2. K-Nearest Neighbors (KNN)
3. Support Vector Machines (SVM)
4. Random Forest
5. XG Boost

9. CHAPTER SCHEME FOR PROJECT REPORT

10. WORK PLAN

Milestone Activities	Milestone Dates
Submission of Synopsis to Guide	01-06-20201
Synopsis discussion with guide	04-06-2021
Final Synopsis submission after discussion with guide to AIMA	10-06-2021
Discussion of progress with Guide	30-06-2021
Interim progress review with Guide	10-07-2021
Project Report draft submission to Guide	20-07-2021
Submission of final Project Report to AIMA	30-07-2021