Forest Cover Type Prediction

Revision Number: 2.0

Last date of revision: 31/08/2023

Document Version Control

Date Issued	Version	Description	Author
01/07/2023	1	Initial HLD — V1.0	Lakshmi Gundala
31/08/2023	2	Updated KPI — V1.1	Lakshmi Gundala

Contents

Abstract

The study area includes four wilderness areas located in National Forest, and each observation is a 30m x 30m patch. We are going to predict the predominant kind of tree cover of each patch from raw form data[2], which contains binary columns of qualitative independent variables such as wilderness areas and soil type. We will use the training set to set the model, and predict an integer classification for the forest cover type of each line in test set, and the seven types are: 1 - Spruce/Fir; 2 - Lodgepole Pine; 3 - Ponderosa Pine; 4 - Cottonwood/Willow; 5 – Aspen; 6 - Douglas-fir; 7 – Krummholz

1 Introduction

1.1 Why this High-Level Design Document?

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

- Present all of the design aspects and define them in detail
- Describe the user interface being implemented
- Describe the hardware and software interfaces
- Describe the performance requirements
- Include design features and the architecture of the project
- List and describe the non-functional attributes like: o Security o Reliability o Maintainability o Portability o Reusability o Application compatibility o Resource utilization o Serviceability

1.2 Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

1.3 Definitions

Forest composition is a valuable aspect of managing the health and vitality of our wilderness areas. Classifying cover type can help further research regarding forest fire susceptibility and de/re-forestation concerns.

2 General Description

2.1 Product Perspective

The forest cover type prediction is a prediction which is used to predict the predominant type of tree in sections of wooded areas.

2.2 Problem statement

Forest land is highly required for developing ecosystem management. Any changes that occur in ecosystem should be carefully noticed to avoid further loss. This model is helpful in noticing the changes occurred due to heavy floods or any other calamities which affected the forest land.

The goal is to predict seven different cover types in four different wilderness areas of the Roosevelt National Forest of Northern Colorado with the best accuracy

Four wilderness areas are:

1: Rawah
2: Neota
3: Comanche Peak
4: Cache la Poudre

Seven categories numbered from 1 to 7 in the Cover_Type column, to be classified:

1: Spruce/Fir
2: Lodgepole Pine
3: Ponderosa Pine
4: Cottonwood/Willow
5: Aspen
6: Douglas-fir
7: Krummholz

2.3 PROPOSED SOLUTION

The solution proposed here is an Forest cover type prediction can be implemented to perform above mention use cases .In first case the forest cover type will predict all the predominant type of trees in the wild area sections.

2.4 FURTHER IMPROVEMENTS

Forest cover type prediction will be more usable to predict the predominant type of tree in section of wild areas and through this we can easily predict the trees and its status. Through this we can understand the forest composition and vitality of wild areas.

2.5 Technical Requirements

This Document address to predict seven different cover types in four different wilderness areas of the

Roosevelt National Forest of Northern Colorado with the best accuracy

Four wilderness areas are:

1: Rawah

2: Neota

3: Comanche Peak

4: Cache la Poudre

Seven categories numbered from 1 to 7 in the Cover_Type column, to be classified:

1: Spruce/Fir

2: Lodgepole Pine

3: Ponderosa Pine

4: Cottonwood/Willow

5: Aspen

6: Douglas-fir

7: Krummholz

2.6 Data Requirements

Data requirement completely depend on our problem statement.

Name	Unit of measurement	Description
Elevation	meters	Elevation
Aspect	degrees azimuth	Aspect
Slope	degrees	Slope
Horizontal_Distance_To_Hydrology	meters	Horz Dist to nearest surface
Hillshade_9am	(0 to 255 index)	Hillshade index at 9am
Soil_Type	(40 binary columns, 0 = absence or 1 = presence)	Soil Type designation
Wilderness_Area	(4binary columns,0=absence or1=presence)	Wilderness area
Cover_Type	(7 types, integers 1 to 7)	Forest Cover Type designation

Vertical_Distance_To_Hydrology	meters	Vert Dist to nearest surface water features

2.7 Tools used

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, TensorFlow, Keras and Roboflow are used to build the whole model.

- Google collab is used
- For visualization of the plots, Matplotlib, Seaborn and Plotly are used.

- AWS is used for deployment of the model.
- Tableau/Power BI is used for dashboard creation.
- MySQL/MongoDB is used to retrieve, insert, delete, and update the database.
- Front end development is done using HTML/CSS • Python Django is used for backend development.
- GitHub is used as version control system.

2.7.1 Hardware Requirements

- Operating system : windows 10
- Processor : intel core i5
- Disk space : 2g

2.8 Constraints

The Forest cover type prediction model is must be user friendly, as it work extremely good to predict the wild area tress and its environment as well its conditions.

2.9 Assumptions

The main objective of this project is to predict seven different cover types in four different wilderness areas of the

Roosevelt National Forest of Northern Colorado with the best accuracy

Four wilderness areas are:

1: Rawah

2: Neota

3: Comanche Peak

4: Cache la Poudre

Seven categories numbered from 1 to 7 in the Cover_Type column, to be classified:

1: Spruce/Fir

2: Lodgepole Pine

3: Ponderosa Pine

4: Cottonwood/Willow

5: Aspen

6: Douglas-fir

7: Krummholz

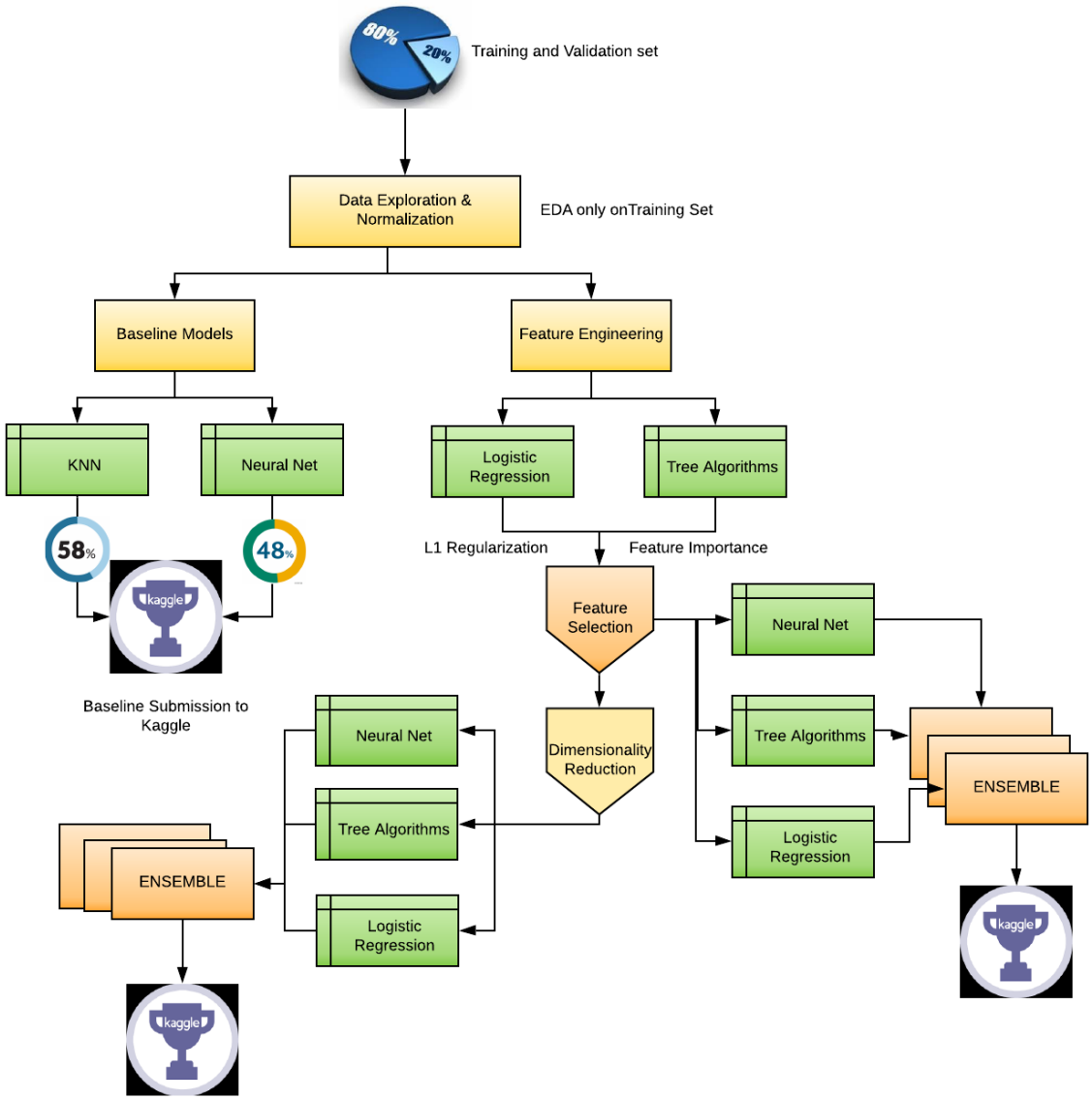
.

3.Design Details

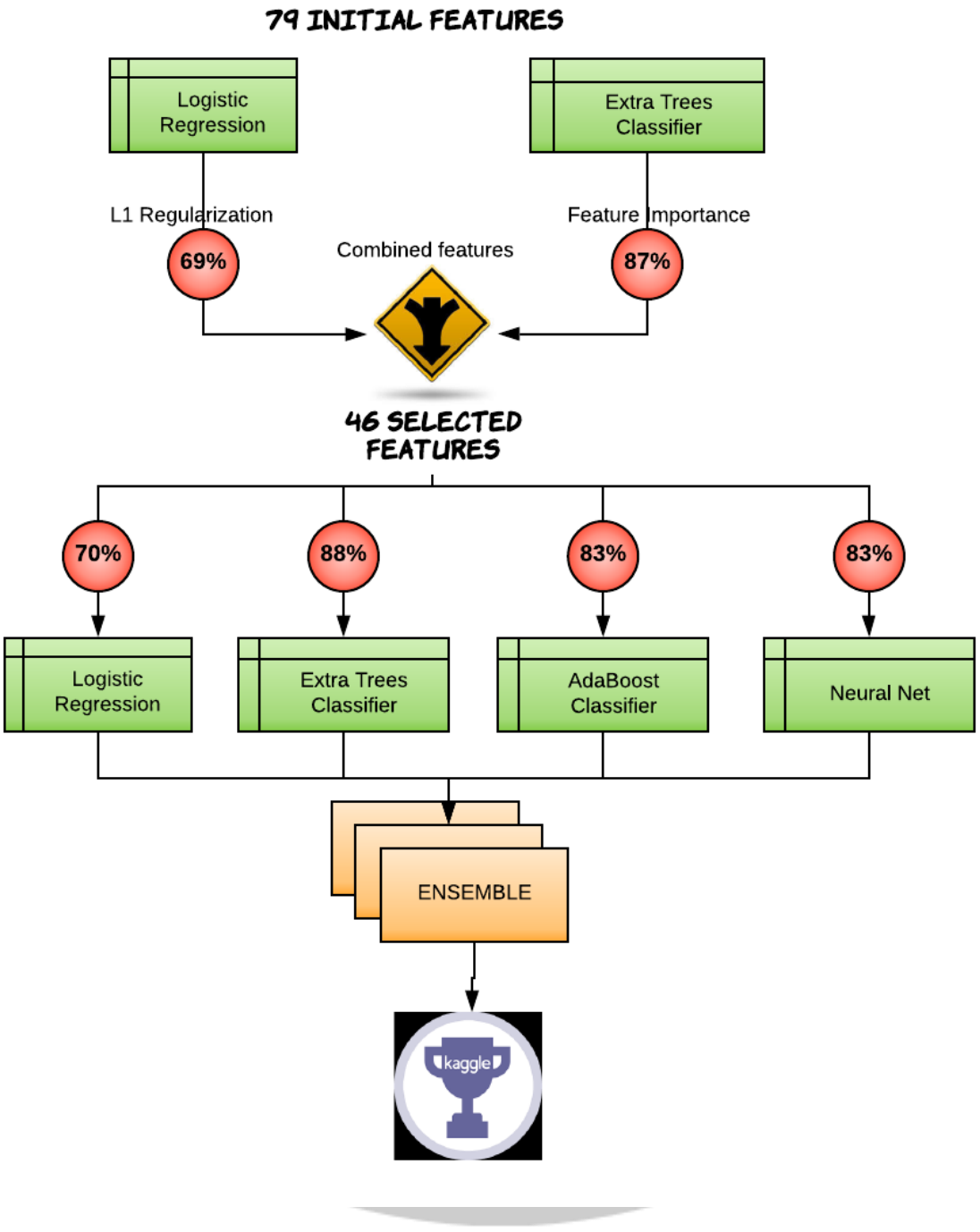
3.1 Process Flow

For identifying the different types of anomalies, we will use a deep learning base model. Below is the process flow diagram is as shown below.

Proposed methodology



3.1.1 Model Training and Evaluation



3.2 Event log

The system should log every event so that the user will know what process is running internally.

Initial Step-By-Step Description:

1. First we will import the packages which are required.
2. Then we will give the dataset as an input and system will read all the inputs and will describe then will show the features of the dataset.
3. After that we will choose the type of visualizations we need to examine the data clearly which is easy to understand.
4. Finally we will represent it in a model like deep learning models.

3.3 Error Handling

Should errors be encountered, an explanation will be displayed as to what went wrong? An error will be defined as anything that falls outside the normal and intended usage.

4.Performance

4.1 Reusability

The code written and the components used should have the ability to be reused with no problems.

4.2 Application Compatibility

The different components for this project will be using Python as an interface between them. Each component will have its own task to perform, and it is the job of the Python to ensure proper transfer of information.

4.3 Resource Utilization

When any task is performed, it will likely use all the processing power available until that function is finished.

4.4 Deployment

Conclusion

Designed the best forest cover type prediction model which predicted seven different cover types in four different wilderness areas of the

Roosevelt National Forest of Northern Colorado with the best accuracy

Four wilderness areas are:

1: Rawah

2: Neota

3: Comanche Peak

4: Cache la Poudre

Seven categories numbered from 1 to 7 in the Cover_Type column, to be classified:

1: Spruce/Fir

2: Lodgepole Pine

3: Ponderosa Pine

4: Cottonwood/Willow

5: Aspen

6: Douglas-fir

7: Krummholz successfully. This model can calculate the accuracy of the wild area trees and its atmosphere and the whether conditions.

.

References

1. https://en.wikipedia.org/wiki/Random_forest
2. https://www.tandfonline.com/doi/full/10.1080/08839514.2020.1771523
3. https://in.images.search.yahoo.com/search/images;_ylt=AwrProJruOFk4S4yoUC7HAx.;_ylu=Y29sbwNzZzMEcG9zAzEEdnRpZAMEc2VjA3BpdnM-?p=images+of+forest+cover+type&fr2=piv-web&type=E210IN885G0&fr=mcafee