Summer Internship Report

Submitted

by

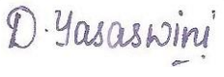
Yasaswini Desu – AP19110010008

Department of Computer Science and Engineering

SRM University-AP, Andhra Pradesh, India

July - 2021

DATA SHEET

Roll Number	:	AP19110010008
Name of the student	:	Yasaswini Desu
Branch & Section	:	CSE D
Batch	:	2019-2021
Type of internship	:	Industrial Training
Company Name/Institute Name	:	Hex N Bit (Tevatron Technologies)
Company/Institute Website	:	https://www.hexnbit.com/
Start Date (MM/DD/YYYY)	:	06/05/2021
End Date (MM/DD/YYYY)	:	07/15/2021
Duration (No. of days)	:	45 days
Status of the internship	:	Completed
Name of internship mentor	:	Dr. Amit Kumar Mandal(amitkumar.m@srmap.edu.in)
GitHub Repository		https://github.com/Yasaswini-3004/Vehicle-Classification-using-ML
Signature of the student	:	

ACKNOWLEDGEMENT

The satisfaction that accompanies that the successful completion of any task would be incomplete without the mention of people whose ceaseless corporation made it possible, whose constant guidance and encouragement crown all efforts with success.

We are grateful to our training and project mentor **Mr. Gagan Preet Singh** sir, R&D Head at Hex N Bit EdTech Pvt Ltd for the guidance, inspiration and constructive suggestions that helped us in the completion of this project.

At the outset we thank our Head of the Department, **Dr. T. Ragunathan** sir for the moral support and excellent facilities provided. I would like to thank all the teaching and non-teaching staff members of Computer Science department who extended their full cooperation during this pandemic for letting us know about this internship.

I also thank all my friends who helped in sharing knowledge with me.

TABLE OF CONTENTS

Sl.No.	Content	Page No.
1.	Introduction	5
2.	Objective of the Internship	6
3.	Skills acquired through the internship	7
4.	Overview of the project carried out during internship	9
5.	Results	11
6.	Conclusion	12
7.	References	12

INTRODUCTION

This report is a short description of my 45 days internship at **Hex N Bit** which was carried out as a compulsory component of Summer Internship. This is a virtual internship which was started on 5th June, 2021 and ended on 15th July, 2021. Since I am interested in Artificial Intelligence and Machine Learning, the work was concentrated related to AIML.

At the beginning of the internship, I formulated several learning goals, which I wanted to achieve:

- Complete grasp on Python Programming.
- Understanding various modules in Python.
- Data Visualization.
- Machine Learning Techniques.
- Open CV

OBJECTIVE OF THE INTERNSHIP

As a part of fulfilments of the requirements of summer internship program at Hex N Bit, I considered AIML as my core.

The primary objective of the internship to generate a through understanding of Python programming, AI and ML and engaging with the trainer and mentor. This would help me to pave a way towards growth in my academic as well as personal development. Apart from the general objectives, the specific objectives are:

- To acquire exposure with the companies.
- Maintaining professional relationship with the mentors.
- To learn and apply theoretical knowledge.
- To develop impersonal, managerial and communication skills.
- To come up with the possible strategies to gain competitive advantage.
- To be a valuable asset for the organization by contributing positive aspects.

The Technical objective of this internship is, to prepare and process the data, exploring the data with visual analysis from that data, preparing and applying the various Machine Learning (ML) models, training, testing and evaluating the prepared ML models. With that experience making us (students) to solve real life scenarios (problems).

SKILLS ACQUIRED THROUGH THE INTERNSHIP

Week 1:

- Introduction to AIML
- Introduction to Python
- Getting started with Anaconda and Jupyter
- Python Data Types, Conditional Statements, Loops and Control Statements.
- Python Functions and Lambda functions.
- Map & Filter
- File Handling.
- *Assignment 1*: Basic Operations in Python (15 tasks given)

Week 2:

- NumPy Basics, Operations, Indexing and slicing.
- Pandas Series and Data Frames.
- Pandas Operations, Fixing Missing data, Merging, GroupBy
- Pandas File Reading and writing.
- *Assignment 2*: Questions related to NumPy and Pandas (22 tasks given)

Week 3:

- Introduction to Matplotlib_Pyplot API
- Matplotlib Object Oriented API
- Data Visualization
- Supervised Machine learning techniques & Evaluation Metrics.
- Linear Regression.
- Logistic Regression
- Decision Tree
- *Assignment 3*: Plotting sin and cos graphs, Given a dataset perform exploratory data analysis, using logistic regression model and predicting, performing metrics.

Week 4:

- Introduction to Support Vector Machine
- Unsupervised Machine Learning techniques
- K Means Theory and application
- K Means clustering
- Association Rule and Apriori Algorithm
- *Assignment 4*: Given a dataset, performing exploratory data analysis, plotting an elbow plot in order to implement K Means Clustering, performing K Means Clustering, Creating a scatter plot as per the clustered values

Week 5:

- Introduction to Computer Vision
- Drawing functions and basic operations in CV
- K Nearest Neighbour Theory
- Optical Character Recognition using KNN
- Arithmetic Operations in Open CV
- Color Spaces Histogram Thresholding
- Bitwise Operations and Masking
- Image processing basics
- Object Tracking using color
- Object Detection
- *Assignment 5*: Opening Webcam, displaying grayscale feed, detecting face, nose, eyes, smile using Open CV & Mask Detection using Open CV.

Week 6:

- Final Project (Vehicle Classification)

OVERVIEW OF THE PROJECT CARRIED OUT DURING INTERNSHIP

Abstract:

The purpose is to classify a given silhouette as one of four types of vehicles, using a set of features extracted from the silhouette. The vehicle may be viewed from one of many different angles.

Dataset Description:

The features were extracted from the silhouettes by the HIPS (Hierarchical Image Processing System) extension BINATTS, which extracts a combination of scale independent features utilizing both classical moments-based measures such as scaled variance, skewness and kurtosis about the major/minor axes and heuristic measures such as hollows, circularity, rectangularity and compactness.

Four "Corgie" model vehicles were used for the experiment: a **double decker bus, Chevrolet van, Saab 9000 and an Opel Manta 400**. This particular combination of vehicles was chosen with the expectation that the bus, van and either one of the cars would be readily distinguishable, but it would be more difficult to distinguish between the cars

Attributes:

We have 19 columns in the dataset:

Compactness, circularity, distance_circularity, radius_ratio, pr.axis_aspect_ratio, max.length_aspect_ratio, scatter_ratio, elongtedness, pr.axis_rectangularity, max.length_rectangularity, scaled_variance, scaled_variance.1, scaled_radius_of_gyration, scaled_radius_of_gyration_1, skewness_about, skeweness_about.1, skewness_about.2, hollows_ratio, class.

Domain:

Object Detection (Vehicle).

Steps and Tasks:

1. **Data Pre-processing**: We have performed Exploratory data analysis and came to conclusion by understanding different plots and checked if there is any missing data.

2. We have used **heat map** of correlation and to understand the attributes which are highly correlated.
3. And used various models to predict the data.

Classifiers used:

1. **Logistic Regression:**
   Logistic regression is a statistical method for predicting binary classes. predicts the probability of occurrence of a binary event utilizing a logit function.
   How Logistic Regression Algorithm works:
   - Select the best attribute using Attribute Selection Measures (ASM) to split the records.
   - Split the data and train/test the model by importing train_test_split from skit-learn.
   - Train the model using LogisticRegression () and then predict the values.
   - At last, evaluate the data by finding accuracy, recall, precision value etc.
2. **Decision Tree Classifier:**
   A decision tree is a flowchart-like tree structure where an internal node represents feature, the branch represents a decision rule and each leaf node represents the outcome.
   How Decision Tree Classifier Algorithm works:
   - Select the best attribute using Attribute Selection Measures (ASM) to split the records
   - Make that attribute as a decision node and breaks the remaining dataset into smaller chunks.
   - As above split the data and train the model by importing train_test_split from skit-learn.
   - Train the model using DecisionTreeClassifier () and then predict the values.
   - At last, evaluate the data by finding accuracy, recall, precision value etc.
3. **Support Vector Classifier (SVM):**
   It is a set of supervised learning methods used for classification, regression and outlier's detection. The Main objective of SVM is finding a hyperplane that divides a dataset into two classes.
   How Support Vector Classifier Algorithm works:
   - Select the best attribute using Attribute Selection Measures (ASM) to split the records.

- Repeat the data splitting steps as in 1 and 2, and then predict the values using SVC () method.
- Usually, the accuracy with this algorithm is low, so we will hyper parameters to tune the algorithm using GridSearch.
- Then if we use GridSearchCV () algorithm to predict the values, the accuracy will increase to some extent.
- We can consider that as the final accuracy for that model.

4. **K Means:**

   K-Means Clustering is an unsupervised learning algorithm that is used to solve the clustering problems in machine learning or data science. This algorithm groups the unlabelled dataset into different clusters.

   The k-means clustering algorithm mainly performs two tasks:

   - Determines the best value for K centre points or centroids by an iterative process.
   - Assigns each data point to its closest k-centre. Those data points which are near to the particular k-centre, create a cluster. Hence each cluster has datapoints with some commonalities, and it is away from other clusters

   How K Means Classifier Algorithm works:

   - Select the best attribute using Attribute Selection Measures (ASM) to split the records.
   - First, initialize K Means and cluster the data categorize them.
   - Now compare the clustered values and actual values which gives us the accuracy.

RESULTS

I have used 4 classifiers in my final project namely Logistic Regression, Decision Tree Classifier, Support Vector Classifier, K Means Classifier.

After using these four classifiers, I came to a conclusion that, **Logistic Regression**, **Decision Tree Classifier** and **SVM** gives better results when compared to SVM (before tuning) and KMeans

So, we will use either of the above models to predict the silhouette as one of the four types of vehicles.

CONCLUSION

Best Tuning Algorithm: Logistic Regression**,** Decision Tree Classifier and SVM

The accuracy of SVM (Before Tuning) is comparatively low when compared to SVM (After Tuning) and K Means also shows accuracy below 0.5

Correlation among the various features can be obtained easily using a heatmap.

Future scope of improvement in the algorithm:

This dataset deals with the classification of vehicles, we can further develop this by using Open CV and Deep Learning which is one of the easiest possible ways to identify the type of vehicle.

At the end of this training, I'm pretty confident that I can build and deploy industry curated projects and I can take any small-scale life scenario, solve them by AIML models

REFERENCES

https://www.kaggle.com/pritech/vehicle-silhouettes

https://www.javatpoint.com/machine-learning

https://www.python.org/

https://www.hexnbit.com/

https://scikit-learn.org/stable/

https://matplotlib.org/stable/gallery/index.html#

https://www.tableau.com/learn/articles/data-visualization

https://pandas.pydata.org/

https://numpy.org/