**Moupriya Roy**

Hill Cart Road, Siliguri ,West Bengal, 734009

Mobile No:+91-7479114173

E-mail: talktomoupriya25@gmail.com

Passport Number: S5342163

**EXECUTIVE SUMMARY**

**Student of Computer Science and Engineering** currently pursuing 4th Semester. Expertise in understanding complex set of data and put them into simplified model of machine learning. As a coding enthusiast and a student of computer science have experience working in small machine learning projects. Looking forward to apply my experience and abilities to the position of a Full Time Intern in a growth oriented organization and look forward to work in a real time project with professional people.

**SYNOPSIS**

 • Young, energetic and result oriented learner who is looking forward to learn from experienced professionals .

 • Pursued 6 months training program from Coding Ninja on Machine Learning.

**AREAS OF PROFICIENCY**

 • Effective Communication

 • Organization and Planning

 • Team working

 • Problem Solving

 • Taking Lead

**KEY SKILLS**

 • Have good knowledge in **Python Language**.

 • Have experience working with **Python Library** like **beautifulsoup** , **numpy , panda , matplotlib, sci-kit learn, seaborn.**

 • Have Brief Knowledge on how to **check in and check out code from GitHub** and do **branching in gitHub**

 • Well versed in **JAVA** which includes **OOPS concept, File handling , String , Thread , Exception Handling .**

 • Have knowledge in **SQL**.

* Have knowledge on using data visualizations tools such as **Excel, Tabeleau, Qlikview.**

**TECHNICAL PROFICIENCY**

* Python
* Python libraries: Numpy,Panda,Scikit-learn,Seaborn,Matplotlib.
* SQL
* Excel
* Tabeleau
* Qlikview
* Probability/Statistics
* Data cleaning
* Data manipulaton
* Data visualization

**CERTIFICATIONS**

 **• NPTEL Online Certification in “Joy of Computing Using Python”**

 **• NPTEL Online Certification in “Soft Skills”**

 **• Certification of Participation in “TELOS 2019 - the ultimate aim(Annual online Technical Fest of MYWBUT.COM)”**

 **• Certification of Participation in “Project Competition Organised by Tech Management Fest Committee”**

* **Certification of completion in Tech trainning(Solutions Pvt Ltd) in “Industrial trainning program on Python Programm Language”**

**ACADEMIC CREDENTIALS**

 • B Tech in Computer Science , Siliguri Institute of Technology, West Bengal University of Technology,2022

**Detail of Projects**

**Project 1**

**A tinder styled dating web app ( Tech Trainning from MyWbut)**

Technology Skills

Web application,Database creation using Wamp server

Team Size (Self made)

1

Duration

march 2017 – Till Date

Description

 • It is a Tinder Clone desktop application which was made using Tkinter, having the features of login and registration, viewing the list of users, sending a proposal, viewing proposal and viewing matches between the current user and other users. Database is created using Wamp server.’Beautifulsoup’ is used for pulling out datas from Mysql database files. A backend system is used for core apllication logic completing the graphical user interphase (GUI) of application. It evolves around adding users necessary Bio and credentials to fecthing those data and showing data of matches.

**Project 2**

**Gradient Dscent-Boston datastet**

Technology Skills

Logistic Regression , Use of Scikitlearn, Visualization using Matplotlib

Team Size (Self made)

1

Duration

march 2017 – Till Date

Description

**Gradient Descent - Boston Dataset**

Boston dataset is one of the datasets available in ‘Sklearn’. I have

used Training dataset CSV file with X train and Y train data. In this project I have come up with Gradient Descent algorithm and thus predictions for the test dataset given.

**Role:**

1. Code of Gradient Descent for N features and come with predictions.

2. Various combinations of learning rates and number of iterations is tested.

3. Used Feature to get better results.

.

**Project 3**

**Gradient Dscent- Combined Cycle Power Plant**

Technology Skills

Logistic Regression , Use of Scikitlearn, Visualization using Matplotlib

Team Size (Self made)

1

Duration

march 2017 – Till Date

Description

**Gradient Descent - Combined Cycle Power Plant**

Combined Cycle Power Plant dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the power plant was set to work with full load. Features consist of hourly average ambient variables Temperature (T), Ambient Pressure (AP), Relative Humidity (RH) and Exhaust Vacuum (V) to predict the net hourly electrical energy output (EP) of the plant.

Dataset:

1. A Readme file for more details on dataset.

2. A Training dataset csv file with X train and Y train data

3. A X test File to predict and make predictions for this file.

**Role:**

1.Gradient Descent for N features is used as a training algorithm and submit results predicted.

Various combinations of learning rates and number of iterations is tested.

2. Files are in csv format; genfromtxt function in numpy is used to load data from csv file and savetxt function is used to save data into a file.

3.Feature extraction and feature scaling is used to get better results. A csv file with only predictions for X test data.

**Project 4**

**Logistic Regression - Titanic Dataset**

Technology Skills

Logistic Regression , Use of Scikitlearn, Visualization using Matplotlib

Team Size (Self made)

1

Duration

march 2017 – Till Date

Description

#### Titanic dataset is one of the datasets available in sklearn.

#### Dataset:

1. A Training dataset CSV file with X train and Y train data

2. A X test File to predict and make predictions for this file.

**Project 4**

**Implementation of Decision Tree in Python from scratch**

Technology Skills

Decision tree , Use of Scikitlearn, numpy, pandas, Concept of oops

Team Size (Self made)

1

Duration

march 2017 – Till Date

Description

#### Iris dataset is one of the datasets available in sklearn.

#### Dataset:

#### Iris dataset.

The dataset includes three iris species with 50 samples each as well as some properties about each flower. The available columns in this dataset are: **Id**, **SepalLengthCm**, **SepalWidthCm**, **PetalLengthCm**, **PetalWidthCm**, and **Species**.

**Role:**

1. The goal is to code a decision tree classifier/regressor from scratch using just NumPy and Pandas.

2. Gini index, gai ratio ,fit function and prediction fuction etc are addred into the Decision Tree classifier to understand the score better through them

#### 3. Implementation is done cosidering the decision tree for OR .

#### 4. Lastly an Exploratory Data Analysis is made through Sklearn to get a comparison between the two classifier – one is from scratch and another one from sklearn

**Project**

**Text Classification Using Naive Bayes**

Technology Skills

NaïveBayes , Use of Scikitlearn, numpy, pandas, Concept of oops,Visualization using Matplotib

Team Size (Self made)

1

Duration

march 2017 – Till Date

Description

#### “Twenty Newsgroup dataset” is downloaded from here

#### Dataset: [Dataset - http://archive.ics.uci.edu/ml/datasets/Twenty+Newsgroups](Dataset%C2%A0-%20http%3A/archive.ics.uci.edu/ml/datasets/Twenty%2BNewsgroups)

**Role:**

1. The goal is to code a Naïve Bayes classifier from scratch using just NumPy and Pandas.
* Frequency count fuction *creates a dictionary which stores several dictionaries corresponding to each news category. Each value in the subdictionary represents the frequency of the key corresponding to that news category and classes represents the different news categories*
* Fit fuction to fit the model i.e train the model i.e train data
* Probaility fuction creates a numpy array counting classes and test data points of model
* Then each best point is predicted out using result of predict fuction.
* Score function returns the mean accuracy of predicted new data points.
* A bar graph is made out to visualize how “Number of Frequency” goes with “Words in vocabulary”

#### Lastly an Exploratory Data Analysis is made through Sklearn to get a comparison between the two classifier – one is from scratch and another one from sklearn.

#### Classification Reports are created for results from both the classifiers.

#### Project Link: <https://www.kaggle.com/moupriyaroy/beginner-text-classification>