Vivek KUMAR Dubey



An Aspiring Data Analyst/Big Data Engineer seeking opportunity to workingloballycompetitiveenvironmentonchallengingprojects, contributingtothesuccessofthe organizationandatthesametimeenhancing my technical skills, leadershipskills and communicationskills, and to seek professional and personal growth.

**CERTIFICATIONS**

**# Core Java Intern** at **Hewlett Packard**, Lucknow, Uttar Pradesh, India, June 2014-July 2014.



Dubey.vivek116@gm

8527498055

linkedin.com/in/dubey.vivek116

**Mini Projects**

**#Predicting the Survival of Titanic Passengers**

The titanic and titanic data frames describe the survival status of individual passengers on the Titanic. Performed **EDA**on the given dataset. Used **Logistic Regression** algorithm to predict survival of a passenger. Achieved **f1-score** of **0.82** for the prediction model.

**#Loan-Prediction**

Used **Ensemble learning** to predict whether a loan application should be passed or not. The system approves or rejects the loan applications.**Used several ensemble** algorithms and improved the accuracy of the model from **68% (XGBoost classifier)** to **75% (Random Forest classifier)**.

**#Diabetes prediction**

Used **Naïve Bayes Classifier Algorithm** to predict whether a person has diabetes or not using parameters like BMI, insulin level, age and so on. Built a basic NB classifier and achieved an **accuracy of 77%**. **Further improved** the accuracy of the model **to 81%** using **k-fold cross validation**method.

**# Temperature Forecast**

Used **Random Forest Regression model** to predict the max temperature for tomorrow in a city using one year of past. Achieved an accuracy of greater than **90%**with **9 features initially**.Done**feature selection** tofind important features and **got almost same accuracy with 2 features, indicating** that **we can achieve** nearly **same performance** by obtaining **only 2 features**, saving resources during data acquisition.

## **#Credit Card Fraud Detection**

Used **Support Vector Machine Classifier, XGBoost classifier** to predict fraudulent transaction. Used **Hyperparameter Optimization** using **GridSearch** to improve sensitivity **from 89% and 91%to 92% and 92%** respectively for each classifier.

**Major Projects**

**#FacialEmotion, Age and GenderRecognitionSystem**

Developed a Computer Vision module for detecting the emotion**(94% accuracy)**, age and gender**(74% accuracy)** of a person in a real-time webcam using**Convolutional Neural Network (CNN) based architecture, Kerasand OpenCV**.

**EDUCATION**

**# Post Graduate Diploma in Big Data Analytics** from CDAC ACTS-Knowledge Park, Bangalore-2020-2021.

**# Bachelor of Technology(B.T)in Information Technology from Dr.MCSCET**at UPTU -2010-2014.

**# 12th Standard [CBSE] –**Smt.DDTIC, GORAKHPUR

**# 10th Standard [CBSE] –**CENTAL ACADEMY GORAKHPUR

**TECHNOLOGY**

**Experience**

**# Implementation Analyst at Mercer, Gurgaon**from October **22, 2015**to **June 25, 2018**.

Creating weekly, monthly, and yearly reporting after fetching data from the different tools.

* Involved in Defect Coordination and tracking of defects to decide on severity & validity of the logged defects in Connect.
* Coordinated with Track leads from state side.
* Prepared and sent Daily Status report including the Execution Summary, Defect Summary reports to the entire team & the Onshore Client.
* Creating VBA tools for on-shore team members.
* Developing MIS for the Team effort and provide the daily basis work alignment by MIS Inhouse created tool.
* the Execution Summary, Defect Summary reports to the entire team & the Onshore Client.
* Creating VBA tools for on-shore team members.
* Developing MIS for the Team effort and provide the daily basis work alignment by MIS Inhouse created tool.

**Skills**

**#** Skilled in programming language **Python, R, VBA**and **Java**.

**# Big Data Technologies** including **Hadoop, Hive, HBase, Kafka** and **Spark**

**#RDBMS** and skilled in **SQL. Basic AWS and Cloud computing.**

**# Machine Learning** using libraries **Scikit learn, NumPy, Pandas, Matplotlib, TensorFlow.**

**#** Visualization tools - **Power BI**

**#**Operating Systems -**Windows** and**Linux**

**# MS Office, MS Excel, VBA**