Gor Vishal Rajeshkumar

Email: rajgorvishal28@gmail.com

Skills:

- Programming languages: **Python** (good proficiency)
- Proficient in **Hadoop framework** with a good understanding of its core components, including **HDFS, MapReduce, and YARN**. Completed **Big data certification** to deepen knowledge of Hadoop ecosystem tools such as **Hive**.
- Knowledgeable in distributed computing systems and **big data** processing platforms, such as **Databricks** and **Apache Spark**
- Familiar with **PySpark** DataFrame operations, including reading and writing data, transforming data, and splitting data into training and test sets
- Hands-on experience in **web scraping** using Python and libraries such as **Beautiful Soup**
- Database management: **MySQL** (good Proficiency), Firebase
- **ETL** processing
- **Data Modeling**
- Operating systems: **UNIX**, Kali Linux, Windows
- Software development methodologies: SDLC, STLC
- Data Visualization: **Power BI, Basic Excel**
- **OOPS** Concepts
- Database Systems: **DBMS, RDBMS**

Projects:

HDFS File Management Project

- Connected to a server with **HDFS** installed and a local system
- Manipulated files and folders on the local system and HDFS using **CLI commands**
- **Transferred files between the local system and HDFS**
- Learned about the challenges of **working with distributed file systems**
- Developed **problem-solving and troubleshooting skills**
- Worked independently and as part of a team, with attention to detail and accuracy.

Data Extraction and Analysis of Amazon's Men's Shoes Category using Python and MySQL

- Scraped men's shoes data from Amazon using **Python's Beautiful Soup** library
- Cleaned and stored data in a **Pandas DataFrame**, exported as a CSV file
- Analyzed data using **MySQL Workbench and SQL queries** to find top 10 most expensive shoes, calculate average prices by brand, determine total number of shoes available per brand, and find minimum and maximum prices per brand
- Analyzed shoe price distribution by range using SQL
- Used SQL **analytical functions** to extract top 3 most expensive and cheapest shoes for each brand
- Demonstrated proficiency in **data extraction, cleaning, analysis, and SQL**.

Linear Regression Model using Databricks and pyspark

- Created a new cluster on **Databricks** and **imported a large dataset**
- Built a linear regression model using **PySpark's ML library** and **handled categorical values** with **StringIndexer**
- Grouped independent columns together with **VectorAssembler** and split the data into training and test sets
- Trained the model on the training set and evaluated its performance on the test set using **evaluate()**
- **Made predictions** using the trained model and calculated evaluation metrics to assess the model performance.

World Population Dashboard

- Data was first collected from the web using the **BeautifulSoup** library and **Pandas**, and then **exported to a CSV file**.
- Created a sample dashboard using **Power BI to visualize the world's population** by 2022 and 2023.
- Designed the dashboard with various data visualizations including **stacked bar charts, pie charts, slicers, donut charts, cards (KPI), a treemap, and tables** to compare data from various countries.
- Demonstrated **proficiency in web scraping, data analysis, and data visualization.**

Vwall – wallpaper app

- **Used API** and Firestore (Database by google) for fetching images
- Firestore was also used so that users can upload there own images
- Implemented unique feature of automatically change wallpaper of home screen in desired timeout entered by user

Education:

- Bachelor of Science in Information Technology (BScIT), KV Pendharkar College • Graduated with a **GPA of 9.8** (out of 10.0)

Certifications:

Big Data, Techlamp Institutions, May 2023