Harsh Kumar

Mumbai   harshkumar707995@gmail.com   7488942448

in www.linkedin.com/in/harsh-sharma-1a051222b

Education

Centre for Development of Advanced Computing (CDAC)   *Agu 2024 – Feb 2025*
PG Diploma in Big Data Analytics

- **Coursework:** Linux Programming and Cloud Computing, Data Collection and DBMS, Python and R Programming Advance Analytics using Statistics,Big Data Technologies, Practical Machine Learning, Data Visualization - Analysis and Reporting ,Object Oriented Programming with Java 8

Veer Madho Singh Bhandari Uttarakhand Technical University   *2020 -2024*
Computer Science and Engineering (CSE)

- CGPA: 7.2
- **Coursework:**

Cooperative public School , Gaya
XIIth

- 70.4

DAV public school , Gaya
Xth

- CGPA: 7.6

Projects

H-1B Visa Status Predictor

- The H1B visa application process is complex and uncertain for both employers and applicants. This project utilizes machine learning to predict visa approval probabilities and extract key insights from historical H1B data. By identifying crucial factors influencing approvals ,it empowers stakeholders with data driven decision making, reducing uncertainty and enhancing application strategies.
- Python, Pandas, NumPy, Machine Learning, PySpark ,Matplotlib, Seaborn,Power BI ,Streamlit, Azure

Big Data Insights: Decoding NYC Parking Violations

- Analyzed NYC parking violations data using Apache Spark and Hive SQL to uncover patterns and trends, optimizing data processing and analytics for large datasets.
- PySpark , Hive SQL

Email Spam Detection Using Machine Learning

- Developed a machine learning model to classify emails as spam or non-spam using various techniques and evaluated its accuracy.
- Tools Used: python , pandas , numpy , matplotlib ,nltk, machine learning , data visualization

Analysis of IMDB Dataset Using SQL and Power BI

- Analyzed IMDB data to extract insights on movie ratings,genres, and trends etc. using SQL and visualized the findings with interactive Power BI dashboards.
-

Technologies

Primary Skills: Python, MySQL

Programming Languages: PySpark, Hive ,java , R ,Shell Scripting ,C ,C++

Data Visualization: Power BI, Tableau, Matplotlib, Seaborn ,Excel

Relevant Skills: Machine Learning, Natural Language Processing, Probability and Statistics, Data Analytics ,ETL , Data Structures, DBMS, Big Data Technologies, Linux , Databricks , Excel .

Cloud : AWS , Microsoft Azure