HARSHAL KAKAIYA

Guelph, ON, Canada (437) 661-4803 harshalpkakaiya@gmail.com
linkedin.com/in/harshal-kakaiya github.com/harshalk612 harshalk.netlify.app

SUMMARY

Detail-oriented Data Scientist with hands-on experience in data analysis, machine learning, and statistical modeling gained through internships and diverse projects. Proficient in Python, SQL, Matplotlib, and Seaborn, with a proven ability to develop predictive models and extract meaningful insights from complex datasets.

TECHNICAL SKILLS

Data Science: Python, R, Data Visualization, Feature Engineering, Machine Learning, Deep Learning
Mathematics for ML & DL: Linear Algebra, Calculus, Statistics, Probability
Python Packages/Frameworks: Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn, Tensorflow, Keras
Web Development: HTML, CSS, Javascript, Django
Databases: MySQL, MongoDB
Cloud Deployment & Containers: Netlify, Git/Github

EDUCATION

Master of Data Science **Sep. 2024 – Sep 2025** ***(Exp.)***
University of Guelph *Guelph, ON, Canada*
- Current GPA: 4.0/4.0

Bachelors of Engineering in Information Technology **June 2019 – July 2023**
Sarvajanik College of Engineering & Technology *Surat, GJ, India*
- Cummulative GPA: 3.7/4.0

WORK EXPERIENCE

Machine Learning Intern **Jan 2023 – June 2023**
Tops Technologies *Surat, GJ, India*
- Engineered a Cancer Classifier project with the team by leveraging advanced machine learning algorithms to predict cancer types with over 90% accuracy, enhancing the reliability of diagnostic tools for healthcare providers.
- Optimized machine learning pipelines by designing and implementing robust data preprocessing workflows and feature selection strategies, significantly boosting model training efficiency.
- Spearheaded the deployment of machine learning models into production by collaborating with data scientists and software engineers, streamlining processes to enable actionable business insights.
- Interpreted and visualized complex datasets to uncover key patterns, driving improvements in model accuracy, precision, and recall for real-world business and other applications.

PROJECTS

Cancer Classifier | *Python, Machine Learning* **April 2023**
- Developed a extensive model to classify cancer types with 91% accuracy, enhancing diagnostic capabilities for early detection
- Conducted data preprocessing and feature engineering, ensuring clean, optimized data for robust model performance
- Evaluated model using key metrics, providing reliable cancer detection to support healthcare professionals in accurate diagnoses

Used Car Price Prediction | *Python, Data Manipulation, Machine Learning* **December 2022**
- Built a machine learning model to predict car prices using features such as mileage, fuel type, and year
- Improved model performance through data cleaning, wrangling, and hyperparameter tuning
- Demonstrated expertise in Python for data processing, model development, and evaluation, achieving strong performance metrics

RFM Analysis for Customer Segmentation | *Big Data, Python, Data Visualization, Clustering* **June 2022**
- Segmented business customers based on Recency, Frequency, and Monetary (RFM) criteria to better understand consumer behavior.
- Utilized Python and libraries like Seaborn, Matplotlib, and Squarify to visualize key insights and trends.
- Developed tailored strategies for different customer segments, optimizing business approaches based on data findings.

LEADERSHIP / EXTRACURRICULAR

Master of Data Science **Fall 2024 – Present**
Student Representative *University of Guelph*
- Acted as a liaison between students and faculty, addressing concerns and advocating for program improvements.
- Represented student interests, ensuring their needs were communicated effectively to faculty and administration.
- Promoted academic growth by supporting initiatives that enhanced the learning experience and fostered collaboration.