

# Harshal Kakaiya

Guelph, ON, N1E 2M4 | (437) 661-4803 | [harshalpkakaiya@gmail.com](mailto:harshalpkakaiya@gmail.com)  
[linkedin.com/in/harshal-kakaiya](https://www.linkedin.com/in/harshal-kakaiya) | [github.com/harshalk612](https://github.com/harshalk612) | [harshalk.netlify.app](https://harshalk.netlify.app)

## Professional 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

**Programming Tools:** Python, R, SQL

**Data Science:** 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

## Projects

**Housing Data Price Prediction** | *Python, Machine Learning, Regression Problem*

**October 2024**

- Executed a detailed machine learning workflow on AMES dataset, including data analysis & preprocessing, to predict house prices.
- Identified Ridge Regression as the top model, achieving a lowest RMSE of 18,407 through effective hyperparameter tuning.
- Analyzed the bias-variance trade-off in Ridge Regression by plotting a U-shaped curve to deepen model performance insights.

**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) 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.

## Internship 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 95% 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.

## Education

**Master of Data Science**

**Sep. 2024 – Present**

*University of Guelph*

*Guelph, ON, Canada*

- Current GPA: 4.0/4.0 with Expected Completion of September 2025
- Relevant Coursework: Introduction to Data Science, Data Manipulation & Visualization, Analysis of Big Data, Machine Learning for Sequences, Analysis of Spatial-Temporal Data

**Bachelors of Engineering in Information Technology**

**June 2019 – July 2023**

*Sarvajanik College of Engineering & Technology*

*Surat, GJ, India*

- Cummulative GPA: 3.7/4.0
- Relevant Coursework: Probability and Statistics, Data Structures, Database Management Systems, Analysis and Design of Algorithms, Artificial Intelligence, Pattern Recognition, Data Compression, Software Engineering

## Leadership / Extracurricular

**Student Representative**

**Fall 2024 – Present**

*Master of Data Science*

*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.