

EDUCATION

Master of Data Science

University of British Columbia (CGPA: 3.8/4.0)

Sep 2024 – Jun 2025

Vancouver, Canada

Bachelor of Technology in Computer Engineering

K. J. Somaiya Institute of Technology (CGPA: 3.8/4.0)

Aug 2020 – May 2024

Mumbai, India

Honours Degree in Data Science

K. J. Somaiya Institute of Technology (CGPA: 3.7/4.0)

Aug 2022 – May 2024

Mumbai, India

WORK EXPERIENCE

Data Science Intern

Heron Law Offices

Apr 2025 – Jun 2025

Vancouver, Canada

- Developed an interactive dashboard using Streamlit and Plotly to visualize immigration inadmissibility and litigation data from IRCC, enhancing transparency and accountability in government decision-making
- Conducted exploratory data analysis and statistical testing using R to examine potential biases and trends in litigations.
- Engineered a robust NLP pipeline combining rule-based methods and LLaMA 3 for automated extraction and classification of metadata from unstructured Canadian Federal Court decisions, achieving 85% accuracy.
- Collaborated closely with legal experts to translate complex legal and data science concepts into clear, user-friendly visualizations and reports for non-technical stakeholders.

Research Intern

SIRAC - Somaiya Institute for Research & Consultancy

Jan 2023 – Jul 2024

Mumbai, India

- Collaborated with cross-functional teams to develop a custom dataset of 75,000 farm plots using OpenCV and stored on Amazon DynamoDB with associated farmer information.
- Developed a deep learning framework based on YOLOv8 architecture using PyTorch to automate the extraction of agricultural field boundaries, digitizing over 100,000 farm plots.
- Reduced farm boundary mapping time and workforce by 50% through automation, improving operational efficiency.
- Built and deployed a fully functional Android application using Flutter enabling 92 field supervisors to independently map farm plots, significantly reducing reliance on technical support.
- Authored a review paper on prior algorithms for agricultural field boundary extraction, currently under review.

PROJECT

Ensemble Rainfall Prediction Model

University of British Columbia

Mar 2025

Vancouver, Canada

- Developed an ensemble machine learning model for daily rainfall prediction in Australia using large-scale climate datasets, employing Python for data preprocessing and feature engineering, achieving an R^2 score of 84.88%.
- Utilized AWS cloud services (EC2, EMR Spark, S3) to manage scalable data processing, model training, and deployment, including setting up a Flask-based REST API for real-time prediction.

Ads Analytics

University of British Columbia

Feb 2025

Vancouver, Canada

- Developed and deployed an interactive web dashboard using Dash, Altair, and Flask for analyzing Google Play Store apps, enabling advertisers to assess app engagement and target high-performing apps.
- Implemented visual analytics features to filter, compare, and identify apps based on installs, ratings, and categories; integrated caching and server deployment with Gunicorn for production-ready performance.

Datpro

University of British Columbia

Jan 2025

Vancouver, Canada

- Developed a modular Python package for data cleaning, anomaly detection (missing values, outliers, duplicates), and statistical summarization using IQR and percentile-based methods.
- Automated data visualization workflows altair, generating tailored plots (e.g., histograms, box plots, heatmaps) for exploratory data analysis.
- Integrated CI/CD with GitHub Actions, deployed documentation using ReadTheDocs, and maintained testing with pytest, enhancing code quality and transparency.

Age Group Predictor

University of British Columbia

Dec 2024

Vancouver, Canada

- Developed a reproducible ML pipeline using Docker, Makefile, and pytest, ensuring standardized model training, testing, and reporting workflows for classifying age groups from NHANES survey data.
- Built and evaluated a logistic regression model using scikit-learn and pandas, achieving 73% accuracy; planned future extensions with SVC, KNN, and Naive Bayes for improved generalization.
- Designed and integrated modular components for scalable experimentation and CI-compatible analysis, aligning with MLOps principles and enabling seamless containerized deployment.

+1 (672) 472-6635
Vancouver, Canada
husmail786@gmail.com

Ismail Bhinderwala

linkedin.com/in/ismailbhinder
github.com/ismailbhinder
ismailbhinder.github.io

COMMUNITY INVOLVEMENT

Mentor, CloudML Discovery Series

Azure Developer Community

Jun 2024
Mumbai, India

- Delivered session on core AI, Machine Learning, and Natural Language Processing techniques using Microsoft Azure Cloud, enabling participants to develop and deploy real-world projects.
- Demonstrated practical implementations using ML libraries such as scikit-learn, XGBoost, and PyTorch, and NLP libraries including Hugging Face Transformers, spaCy, and NLTK.

SKILLS

- **Programming Languages:** Python, R, SQL, PySpark
- **Machine Learning & AI:** Supervised Learning, Unsupervised Learning, Deep Learning, Feature Engineering, Optimization
- **Data Processing & Analysis:** Data Cleaning, ETL Processes, Data Visualization
- **Development Tools:** Git, Docker, Flask, Flutter, MongoDB, AWS, Azure, CI/CD, pytest, Poetry, Makefile
- **Libraries** pandas, matplotlib, NetworkX, altair, PyTorch, TensorFlow, NumPy, scikit-learn, tidyverse, ggplot, XGBoost, Plotly
- **Statistics:** Descriptive Statistics, Inferential Statistics, Hypothesis Testing, Probability Distributions, A/B testing
- **Soft Skills:** Critical Thinking, Presentation, Adaptability, Strategic Thinking, Collaboration, Problem-Solving