+1 (437) 557-6009
Canada
harbhajan2105@gmail.com

Harbhajan Singh

Data Scientist

github.com/harbhajan21
scholar.google/harbhajan21
linkedin.com/in/harbhajan21
harbhajan21.github.io/portfolio

TECHNICAL SKILLS

Programming Languages	Python, R, SQL, C, LaTeX
Libraries & Frameworks	Scikit-learn, Tensorflow, Keras, PyTorch, Transformers, YOLO, OpenCV, MediaPipe, dlib, caffe, NLTK, NLP, NER, Data Extraction, SpaCy, ScanPy, Pandas, NumPy, SciPy, Scikit-learn, matplotlib, seaborn, plotly, PySpark, CNN, RNN, Flask, FastAPI, Streamlit, Docker, Kubernetes, SQLite, Linux, MLflow, huggingface
Database	MSSQL, MySQL, Impala, PostgreSQL, MongoDB
Tools	Git, GitHub, Advanced MS-Excel, Talend ETL Tool, CI/CD, UNIX Tools, Shell Scripting, Jupyter NoteBook, RStudio, Anaconda, SSMS, Cloudera, Big Data, Spark, MS-Suite, JIRA, Spyder, VS Code, Google Colab, Azure Power Platform, Prompt FLow, Azure OpenAI Services, Semantic Kernel, Langchain, Embedchain
Visualization Tools	Tableau, Power BI, Google Data Studio
Data Science Skills	Data Analysis, Data Mining, Statistical Analysis, Predictive Analysis, Time Series Forecasting, Statistics, Recommendations System, Machine Learning, Deep Learning, Computer Vision, Artificial Intelligence, Generative AI, Natural Language Processing, Large Language Models, GPT
Cloud Deployment	Google Cloud Platform (GCP), Microsoft Azure, Heroku
Web Development	HTML, CSS
Core Competencies	Analytical and problem-solving, Critical Thinking, Team Leader, Business Acumen, and Communication

EDUCATION

Master of Applied Computing (Specialization in AI), *University of Windsor, Windsor, ON Canada* JAN 2023 - APR 2024

WORK EXPERIENCE

Data Scientist (Part-Time) Sep 2023 - Present
Voiceflip Technologies Inc. *Remote, Ontario*

- Led end-to-end development of a **RAG-based** (Retrieval-Augmented Generation) product for extracting text using **OCR techniques** from diverse PDF formats, including scanned documents, pertaining to **Home Owner Association** (HOA) data.
- Using **BERT** implemented rigorous data correction procedures to ensure accuracy, rectifying spelling errors within extracted text and storing data efficiently in a database.
- Developed a user-friendly **RAG and Streamlit-based UI**, integrating Large Language Models (LLMs) like GPT-3.5-turbo, to provide quick and precise responses to user queries based on the connected database as a knowledge base.

Artificial Intelligence Developer Intern JAN 2024 - APR 2024
Agriculture and Agri-Food Canada - Government of Canada *Harrow, Ontario*

- Worked on a RAG-based chatbot project (AgPal), automated test scripts using **Azure OpenAI** services, including **Prompt Flow**, **Semantic Kernel** and **LangChain**, to enhance system functionality and efficiency.
- Utilized **Prompt Flow** to create LLM-based chatbot/solutions and evaluated the LLM responses by ensuring robust **chatbot performance** based on **GroundTruth** and **Groundedness** score.
- Developed and implemented new features for a deduplication tool, fixing numerous existing bugs to improve tool reliability.
- Conducted data collection for the detection and classification of arthropods and set the data pipelines by collaborating with a research center to build comprehensive datasets.
- Led the Internal Project Management project and increased the overall productivity of the team by 30% by developing a dashboard to monitor various team projects, track project status, and tasks, and assess team capabilities.

Data Analyst MAY 2021 - NOV 2022
TATA CONSULTANCY SERVICES *Bengaluru, India*

- Collaborated with stakeholders and reporting & analytics team at a leading global investment banking and financial services firm (**Credit-Suisse**) to construct and deploy data-driven solutions using Python, SQL, and other tools & software.
- Designed and accomplished **metadata pipelines** by utilizing data lakes and data warehousing concepts to efficiently store and process large data sets by improving performance, empowering data scientists to formulate predictive models and forecasts for customer queries and solutions.
- Led & managed **value-added projects**, including the development of a rule-based Chatbot by utilizing NLP techniques like text mining & topic modeling and the creation of two innovative products i.e. **SQL checklist matching** & **JIRA monitoring**, and achieved high accuracy and efficiency in both products.
- Applied a range of tools and technologies including business intelligence tools like **Tableau & Power BI** and delivered more than 50 dashboards to clients.
- Performed exploratory data analysis (EDA), feature engineering, model training, and model evaluation for a variety of business use cases.
- Executed **Time Series forecasting** to predict the number of requests and reports raised and requested by users & customers also Leveraged AI and ML to produce reports, resulting in an 85% reduction in manual efforts for business analysts and customers.

- Worked on a robust recommendation system using machine learning for personalized financial product recommendations and investment portfolio optimization.
- Enhanced data processing and modeling workflows for 90% efficiency boost, collaborating with cross-functional teams to drive business objectives and tackle intricate challenges. Worked in team with Agile Methodologies and handled scrum calls.

Machine Learning Engineer FEB 2021 - MAY 2021
ARTIFUTECH PVT. LTD. *Gurgaon, India*

- Developed and executed computer vision algorithms for **image processing** and **anomaly detection** leveraging camera feeds.
- Implemented a Face AI application predicting wrinkles, dark circles, age, skin glow, and dark spots on human faces and achieved accuracy up to 93.50% and Deployed **7 web-based AI** and **ML** applications on the **Heroku server** using **CI/CD pipelines**.
- Led end-to-end deployment of deep learning models on cloud platforms, specifically on **Azure** and **Google Cloud**.
- Designed, developed and maintained Machine and deep learning pipelines based on computer vision.

Data Engineer MAY 2020 - JAN 2021
TECHNODATA ANALYTICS SERVICES *Noida, India*

- Produced reports with the help of **SQL Server, MySQL, Talend ETL Tool, PySpark**, and created dashboards to visualize dataset for healthcare and manufacturing clients based on specific requirements. Improved customer feedback and suggestions.
- Proposed and contributed to two value-added projects, including a customer retention program in e-commerce and a web-based duplicate checker tool. Used Python for coding and worked on code review.
- Leveraged advanced machine learning and statistical techniques to design and evaluate algorithms for improving model performance, quality, and accuracy up to 96.25%.
- Utilized **MLFlow** for model serving and monitoring on **Databricks**, ensuring robust performance and reliability of machine learning models in production environments.

Junior Data Scientist MAY 2019 - MAY 2020
CADENTURE INDIA PRIVATE LIMITED *New Delhi, India*

- Implemented and optimized traditional machine learning algorithms (e.g., Random Forest, SVM, KNN) for image classification tasks, improving model accuracy by 15% and reducing processing time by 30%.
- Developed a computer vision pipeline using OpenCV and TensorFlow for real-time object detection, achieving 90% accuracy on a custom dataset of 10,000 images.
- Collaborated on feature engineering and selection processes, utilizing statistical techniques and domain knowledge to enhance model performance for predictive maintenance applications.

Data Scientist Trainee MAY 2018 - MAY 2019
CADENTURE INDIA PRIVATE LIMITED *New Delhi, India*

- Developed and implemented machine learning models using Python and scikit-learn, resulting in a 20% improvement in predictive accuracy for customer churn analysis.
- Collaborated with cross-functional teams to clean, preprocess, and analyze large datasets using SQL and Pandas, contributing to data-driven decision-making processes.
- Assisted in creating and maintaining interactive data visualizations using Tableau, effectively communicating complex insights to non-technical stakeholders.

RESEARCH PAPER & PUBLICATIONS

- Dynamically Visualizing Key Performance Indicators for Informed Business Decision-Making (2024) - (Paper Link)
- Customer Segmentation in E-Commerce to Retain and Gain the Customers (2020) - (Paper Link)
- Vulnerability Assessment, Risk, and Challenges Associated with Automated Vehicles Based on AI (2019) - (Paper Link)

PROJECTS

Intelligent Resume Analyzer (Application Link) May 2024 - JUN 2024
Personal Project *Windsor, ON Canada*

- Developed a RAG-based application powered by **Google's Gemini Pro** as an LLM to assist job seekers and recruiters. Features include ATS score calculation, resume analysis with feedback, and customized cover letter generation.
- Successfully deployed the application on **Streamlit cloud**, making both the tool and its source code publicly available. This initiative aims to support job seekers globally by providing free access to advanced resume optimization tools.

Visualizing Sales Data Using Tableau and Python JUN 2023 - JUL 2023
University of Windsor *Windsor, ON Canada*

- Worked on a research project with team members aimed at enhancing the utilization of Key Performance Indicators (KPIs) using the BI tool Tableau. and Ensured data reliability by identifying and addressing outliers using Z-score and Interquartile Range.
- Built interactive Tableau dashboards merging sales data and dynamically responsive KPIs, showcasing real-time insights using data analytics and strengthening data-driven decision-making. (**Dashboard Link**)

A Detailed Study of Different Deep Neural Network Architectures for Traffic Sign Classification **(Link)** MAY 2023 - JUN 2023
University of Windsor *Windsor, ON Canada*

- Experimented and evaluated multiple neural networks architectures (AlexNet, DenseNet, ResNet50, ResNet101, VGGNet, custom CNN) in a team of three for traffic signs classification, reaching **97.31% accuracy** with **custom CNN**.
- Trained neural network model for multi-class classification of traffic signs through extensive training with a large image dataset, employing hyperparameter tuning, early stopping, dropout, pooling layers, categorical cross-entropy loss, and Adam optimizer.

TECHNICAL INVOLVEMENT & EXTRA-CURRICULAR ACTIVITIES

- Achieved second-highest accuracy (96.70%) and **2nd rank** at Future Mobility Challenge Hackathon organized by **St. Clair College, Windsor** by developing a Traffic sign classification solution employing diverse neural network architectures.
- Volunteered at University of Windsor's Orientation Day in May 2023, showcasing strong communication and organizational skills to support new students and Contributed to the success of a CS project demo day event at the University of Windsor.

CERTIFICATES & TECHNICAL TRAINING

- Completed **50+ certifications** in fields of data science, machine learning, and AI from Coursera, Udemy, DataCamp, LinkedIn, Microsoft, Google, etc.