ROHAN J. PATEL

Toronto, ON | +1 (647) 864-9705 | Email: rohan00597@gmail.com | LinkedIn | GitHub | Portfolio **(Open to Relocate)**

SUMMARY

- Experienced professional with 3+ years of experience, with a master's degree in Analytics focusing on Machine Learning.
- Implemented supervised & unsupervised algorithms, Hypothesis & A/B testing to validate performance metrics.
- Designed and executed comprehensive reports & interactive dashboards in Tableau, Power BI, MS Excel & Data Studio.
- Developed, deployed and maintained services in production environments, leveraging AWS, Docker and Kubernetes.
- Adept in relational/non-relational database, JavaScript, API & MVC architecture. Proficient with Git, Jira & Jenkins to perform version control & CI/CD.

TECHNICAL SKILLS

Languages:	Python (Pandas, Numpy, Matplotlib, Scikit-learn, sktime, Seaborn, Plotly), R, Node.js, Javascript, C.
Databases:	MYSQL, NoSQL, MS SQL Server, SQLite, Mongo-DB, PL/SQL, Oracle, Django, HIVE, Influxdb.
Analytics:	Tableau, Power BI, MS Excel, Spark, Hadoop, Grafana, Telegraf, Looker, Vertex AI, AutoML, AWS.
Statistics:	Hypothesis and A/B Testing, Chi-Square, ANOVA, Data Distribution - Bernoulli, Poisson, Exponential.
Tools & IDE:	GIT, Bitbucket, ETL, SPSS, SSRS, Google Colab, Rabbit-MQ, Docker, Pycharm, VS Code, Azure ML.
Big Data:	HDFS, PySpark, Databricks, MapReduce, GCP, SageMaker, LLM, Hugging Face, Langchain, Ollama.
Machine Learning:	Regression, Decision Trees, Computer Vision, XGBoost, LightGBM, K-Means, TensorFlow, Keras.

WORK EXPERIENCE

Data Scientist - AI Training, Contractor | **Outlier**, Canada *Mar 2024 - Present*

- Utilizing advanced Reinforcement Learning with Human Feedback (RLHF) techniques to fine-tune LLM model performance, enhancing the quality of generated code and responses.
- Training LLM models for diverse use cases, incorporating Python, JavaScript, and RLHF techniques to improve prompt accuracy and optimize performance in code generation and NLP tasks.
- Developing and optimising training datasets, collaborating with cross-functional teams to provide feedback on submitted prompts, and improving overall model efficiency and effectiveness in real-world applications.

Data Scientist | Bit Space Development, Manitoba, Canada (Apprenticeship) *Apr 2023 – Jun 2023*

- Curated & annotated an extensive dataset for surgical tool detection, enabling effective YOLO model training, through extensive data augmentation and fine-tuning for improved performance in diverse operating conditions.
- Implemented various YOLO (v5, v6, v8) models to evaluate model performance and identify the best-performing model. Incorporated iterative feedback on the surgical tool detection ensuring suitability for real-world surgical environments.
- Achieved impressive precision and recall in the model by optimizing computer vision & deep learning techniques, resulting in a 20% acc. improvement over the validation set. Leveraged TensorBoard for model visualization & analysis.

Data Scientist, Co-Op | CIBC, Toronto, ON *Sep 2022 – Apr 2023*

- Conducted Time Series Analysis using various Python ML libraries (Prophet, sktime, scikit-learn) to identify trends and forecast new TFSA account creation count resulting in an 87% improvement in the forecast accuracy for campaigns.
- Implemented various ML algorithms pipelines (Logistics, Random Forest, XGBoost, KNN, SVM) to ensure & evaluate data preprocessing, feature engineering, model performance, hyperparameter tuning & identifying the best-performing model. This resulted in a 30% reduction in manual effort for model development.
- Improved monthly forecasted volume for Indirect Auto Lending by employing advanced Time Series Analysis techniques and feature engineering, resulting in significant insight for business growth with an 83% prediction accuracy.
- Developed Python scripts to evaluate time series model performance, collaborating with a colleague to design interactive dashboards and reports utilizing MS PowerPoint, showcasing exceptional teamwork & communication skills.
- Evaluated and improved a web scraper in Python to gather all active bank data (1985-present) and create a separate repository from the OSFI website, for informed strategic decision support with a 98% high success rate.

Software Engineer (NLP), Contractor | Gupshup Technology, Mumbai IN *Mar 2022 – Sep 2022*

- Designed a user-friendly chatbot for Indian Railways enhancing the travel experience for all age groups, resulting in a 25% increase in reservation ticket sales and positive reception from passengers.
- Managed and deployed a Flow BOT by leveraging the Gupshup IDE bot platform and Node.js, integrating it with either MYSQL or Mongo-DB. This resulted in an estimated 40% reduction in development time.
- Created an NLP (Natural Language Processing) demo chatbot with multiple language support to enhance user accessibility, achieving an estimated 85% accuracy in user intent recognition.

Data Scientist | AwanTunai, Jakarta, Indonesia (Apprenticeship) *Jan 2022 – Apr 2022*

- Remodelled apriori algorithm for association mining using ML to identify item association among the 1000's SKU-Ids.
- Increased inventory optimization by 20% by identifying top-selling SKU IDs and buying patterns using time series analysis and classification algorithms.
- Created interactive dashboards using Microsoft Excel & Tableau that improved data-driven decision-making in Fintech companies, resulting in a 28% increase in sales of complementary products (KPI).

Software Engineer (NLP) | Gupshup Technology, Mumbai IN *Apr 2021 – Aug 2021*

- Leveraged strong communication and collaboration skills to bridge the gap between cross-functional teams and clients. Resolved multiple bot issues and **built 15+ chatbots** across diverse platforms for a seamless user experience.
- Administered training programs in JavaScript, Node.js & MySQL, equipping new team members with a solid foundation in coding principles. This leadership experience fostered effective communication & collaboration within the team.

Backend Developer | Credence Analytics, Mumbai IN *Jan 2020 – Jul 2020*

- Designed an interactive real-time dashboard system for monitoring company resources (memory, CPU, logins etc.) and expanded it to monitor microservices & SQL query hits using analytical tools such as pm2, Grafana, InfluxDB, and Telegraf.
- Developed new functionality to improve client-side notifications using RabbitMQ, Increasing customer satisfaction by 10% (KPI) with RabbitMQ-powered real-time client notifications for banking asset management & finance automation.
- Devised and documented an automated test script in Node.js employing super test script, saving 30% time in testing.

EDUCATION

Master of Professional Studies in Analytics (Conc. Applied Machine Intelligence), GPA 3.85/4.00 *Sep 2021 – Jul 2023*
Northeastern University, Toronto, ON
Coursework: Quantitative Analytics, Probability, Statistical Modelling, Data Modelling, Advanced Analytics AI, Big Data and Analytics, Data Structure and Algorithms, Enterprise Analytics, Data Warehouse Predictive models.

Bachelor of Engineering | Major in Information Technology *Jun 2015 – May 2019*
University of Mumbai, India
Coursework: Mathematics, Operating Systems, Cloud Computing, Data Mining, Business Intelligence, DBMS.

RESEARCH PROJECT

Helper Tool [Llama3, Mistral, Langchain, Ollama, Huggingface] *May 2024*

- Developed a standalone local homework helper tool employing the Meta Large Language model 'Llama3' and 'Mistral'.
- Utilized Langchain framework for integrating Language models. Employed Streamlit framework for the user interface.
- Configured Langserve environment for monitoring output, Token count, and Cost (for Paid Model). Integrated Langchain API for chaining prompt and retrieval for component integration.

Smart Cartridge Mask: Dynamic Analysis [Computer Vision, Image processing, OpenCV] *Dec 2023 – Jan 2024*

- Engineered a sophisticated algorithm automating fired bullet cartridge image masking for component identification.
- Applied dynamic geometric computations for precise identification of firearm components such as direction, firing pin drag & breech face, leveraging flood fill and contour-based masking to enhance analysis of fired bullet cartridge cases.

Used Car Price Prediction [R, RStudio, Tableau] *Apr 2023 – Jun 2023*

- Performed robust data cleaning and exploratory data analysis (EDA) using data visualizations to identify trends and relationships between features and target variables for used car price prediction.
- Utilized chi-square testing to identify the attribute's relationship with the target variable. Evaluated Linear regression models with Lasso & Ridge regularization to prevent overfitting, achieving accuracy with an 87% F1 score.

Spotify Music Insights [Python, SQL, Machine Learning, EDA, Excel, PowerBI] *Jan 2022 – April 2022*

- Spearheaded a robust end-to-end data pipeline, integrating ETL processes for data manipulation & feature engineering.
- Applied t-tests to assess musical characteristics & optimized data structures for seamless retrieval and analysis.

Breast Cancer Cell Prediction & Classification [Python, Excel, Jupyter Notebook] *Mar 2022 – May 2022*

- Analyzed the Wisconsin data using classification & clustering models to determine error severity in a confusion Matrix.
- Enhanced SVM with hyperparameter tuning via GridSearchCV, applied PCA for dimensionality reduction, and selected the best model based on accuracy and F1 score.

Eat & Out Recommendation System [JavaScript, HTML, CSS, Django, ORACLE, SQL] *Nov 2018 – May 2019*

- Developed a recommendation system using collaborative filtering and sentiment analysis, publishing findings in IJRESM.