RISHABH KALAI

Canada | rishabh.kalai@outlook.com | LinkedIn | Portfolio

PROFILE

Data Scientist with **nearly 2 years of professional experience**, proficient in **Python, R and SQL**, with a proven track record of cross-functional project management, collaboration and delivering insights.

EDUCATION

University of New Brunswick | Master of Computer Science (**MCSC**) **Sep.2022 – May.2024**
4.0 GPA; Courses: *Applied Statistical Methods in R, Machine Learning, Business Data Analysis, AI, NLP*

BNMIT | Bachelor of Engineering - Computer Science & Engineering (**B.E-CSE**) **Aug.2017 – Jun.2021**
8.35 GPA *(First Class w/ Distinction)*; Courses: *Python Programming, Machine Learning, Big Data Analysis*
Best final-year project presentation award recipient; 2-time Business Case Study Competition Winner

Venkat International Public School (VIPS) | High School (CBSE) **Jun.2002 – Apr.2017**
92% – 12th Grade | **96%** – 10th Grade; *Chief Prefect, Football Team*

WORK EXPERIENCE

University of New Brunswick, *Canada* **| Software Developer,** *Database System* **May.2023 – Apr.2024**

- **Systematically improved processing speed by 40%** for the university's Academic Advisory Tool through code refactoring, testing, redesigning software architecture and parallelized execution.
- **Consolidated data** from multiple sources into a unified, **high-quality data model** using bespoke data wrangling and preprocessing techniques, that facilitated **data cleaning, transformation, and integration.**
- Solely restructured system with new software architecture and design to ensure scalability & compartmentalization. Improved the codebase by **reducing redundancy in the codebase by nearly 30%.**
- Collaborated with academic advisors to introduce statistical analysis and data visualization functionalities, **facilitating strategic decision-making,** and enhancing student success.
- Implemented **k-means clustering** to segment students based on academic performance, and thereby enabled academic advisors to create a more precise and tailored academic intervention for students.

Definitive Healthcare, *India* | **Analyst** **Aug.2021 – Jun.2022**

- Solely developed an **A/B testing module in python for marketing data**, automating data preprocessing, feature selection, experimental setup, data visualization and statistical analysis.
- Integrated a Propensity Score Classification algorithm using logistic regression to achieve **90% accuracy in matching test and control subjects**, significantly improving the reliability of marketing treatment analyses.
- Developed a time-alignment feature to synchronize disparate time-series data, **thereby increasing the precision and coherence** of temporal analysis for client data.
- **Automated the analysis of test experiment using ANCOVA** (Analysis of Covariance) results that allowed the clients to identify the types of customers or markets that respond best, and what marketing treatments to target to **maximize ROI.**

The SmartBridge, *India* | **Machine Learning Intern** **May. – Jun.2020**

- Leveraged multiple regression models to predict life expectancy in a healthcare study, analyzing individual and regional data to enhance prediction accuracy.
- Executed a **comparative analysis study** of classification and regression models to determine the most accurate for predicting life expectancy, evaluating each model's performance against established KPIs.

SKILLS

Python | R | SQL| Excel **| PowerBI | Machine Learning** | Data Science | **Statistics**
Techniques: Classification | Regression| Clustering | Neural Networks | Data Wrangling & Preprocessing
Frameworks: pandas | NumPy | scikit-learn | matplotlib | git | Tidyverse | dplyr
Languages: English (*Fluent – TOEFL 118/120*)

PROJECTS

Classification Algorithms | *Machine Learning* **Jan. – Apr.2024**

- Led a 3-member team on development of custom ML models in Python **from scratch** with **zero reliance on pre-built libraries**, achieving relevant business outcomes in healthcare and consumer datasets.
- Implemented 4 models from scratch: **ANN, Random Forest, AdaBoost & kNN** that achieved on **average 90% accuracy in real-world datasets**, highlighting the effectiveness of the models in diverse applications.
- **Evaluated the model using 5-fold cross validation** in parallelized execution to ensure accurate & efficient evaluation of model performance and reduce overfitting.

Fredericton Transit & Weather | *Data Analysis* **Aug. – Dec.2023**

- Discussed internal data possibilities with the **City of Fredericton's Transit Department**, extracting critical pain points to define project objectives and provide context to the data published publicly.
- Led a 4-member team focusing on **identifying seasonal trends** to plan year-round transit resource allocation; **clustering** to identify and target transit user groups to effectively increase satisfaction.
- Optimized bus schedules for peak and low usage periods, reducing congestion and increasing user satisfaction.

PLUTUS: Personal Finance Manager | *Machine Learning* **Oct.2020 – Jul.2021**

- Designed & developed a standalone desktop-application that encapsulates management of the fundamental aspects of personal finance for an individual. The system consisted of three modules: Expense Manager (**Linear Regression**), Investment Portfolio Manager (**RNN**), and Retirement Planner (**Heuristics**).
- Acknowledged with the **'Best Presentation Award'** for delivering a standout presentation during the final group panel evaluation of the project.

RESEARCH EXPERIENCE

Machine Learning for Predictive Analytics in Personal Finance | CoMSO 2021, Springer **Dec.2021**

Proposed a comprehensive analysis technique that can be utilized to manage the key financial aspects of an individual using machine learning. Three models were implemented: **Linear Regression** for Expenditure Prediction, **RNN** for Investment Portfolio Management and **Logistic Regression** for Retirement Prediction.

Sentiment Analysis to Detect Depression in Social Media Users | ERCICA 2020, Springer **Sep.2020**

Reviewed the different methods that are used in **sentiment analysis** for detecting and diagnosing depression in social media users also proposed a cumulative analysis method to detect the depression level of a person by extracting data from social media posts in conjunction with other data pertaining to the user such as sleep, food intake and exercise patterns.

CERTIFICATIONS

Neural Networks and Deep Learning | deeplearning.ai **Oct. – Nov.2020**
ML & Data Science Using Python | Udemy **Apr. – Jun.2020**
Python Application Programming | IIT Bombay **Apr. – May.2020**
Cloud Computing | IIT Madras **Feb. – Apr.2020**
Advanced Python Bootcamp | Udemy **Jan. – Apr.2020**
Data Science for Engineers | IIT Madras **Jun. – Sep.2019**

CONFERENCES & WORKSHOPS

Micro-Services Architecture Workshop | **Tequed Labs,** *India* **Mar.2020**

- Designed and developed **Java programs** to illustrate the functioning and **purpose of a micro-service** architecture with the concepts of REST API, cloud storage and docker.
- Explored how to plan and manage a mi**gration from a monolith** to the micro-service architecture.
- Gained a high-level understanding of how technical choices in code can **impact the architecture itself**.

RACE 2019 Conference & Hackathon | **REVA,** *India* **Aug.2019**

- Gained a hands-on experience through the workshop on **'Computer-Vision through AI'** that focused on the fundamentals of Neural Networks and how Computer Vision works in real world scenarios. Trained a Convolution Neural Network using Kaggle data to **perform real-time object detection**.

Embedded Systems Workshop | **Tequed Labs**, *India* **Mar.2019**

- Utilised devices and sensors that worked with an **Arduino-Uno Module** and its associated software drivers. Developed programs to provide real-life smart home capabilities with light, heat, and smoke sensors.
- Worked on the **designing and construction of a prototype car** that can be controlled remotely by a programmed script. The device was also designed to output live visual feedback through an attached camera.