

Pushya Jain

Major in Computer Science and Statistics, 4th year

pushya22.jain@gmail.com

+1-604-655-9710

[GitHub/PUSH-YA](#)

[LinkedIn/pushya-jain](#)

Technical skills

Languages/Tools: Python, R, Julia, MATLAB, SQL, Java, TypeScript, C/C++, C#, HTML, CSS, PowerBI

Data Science: Hypothesis testing, ML (feature engineering, A/B testing, CNN, ResNet), NLP (NER, LDA), Computer vision (SIFT, HoG, R-CNN), Time series analysis, Markov Chain Monte Carlo.

Data engineering: PostgreSQL/MySQL, MongoDB, GraphQL, Apache Spark, Hadoop, Airflow, Azure, AWS, ETL pipelines.

Frameworks/Libraries: TensorFlow, PyTorch, Scikit-learn, OpenCV, NLTK, Tidyverse, glmnet, Django, SpringBoot, pySpark, Gym.

Technical Work Experience

Data Engineer and Analyst

Jan 2024 – Apr 2024

First Nations Health Authority, Vancouver

- Created a comprehensive ETL processes to manage data integration processes, enabling real-time data logging and data warehouse updates; the system now handles 100,000+ records weekly with minimal manual intervention.
- Engineered an innovative reporting framework that merged data from 4 different databases; streamlined analysis for the business analysts, which improved data-driven decisions and facilitated a 25% faster reporting cycle.
- Implemented mechanisms for handling data quality issues and conducted data profiling to identify and address errors in Data marts and Data warehouses.
- Researched and planned implementations for adding machine learning to optimise data integration task and improve the business intelligence reporting to the clients.
- **Learned:** SQL, SSMS, SSIS, SSAS, SSRS, ETL, data modelling, data visualisation, PowerBI/Tableau, Azure devops.

Undergraduate Teaching Assistant (UTA)

Sept 2022 – Present

University of British Columbia, Vancouver

- Facilitated classes of up to 60-100 students to increase engagement and understanding of the material through tutorials for Math department (Differential and Integral calculus courses).
- Conducted Computer science labs aiding students to solve complex computation and programming problems for the Computer Science department (Fundamental programming and Introduction to AI course).
- Evaluated assessments and examinations, ensuring a fair and consistent process for both Math and Computer Science departments.
- **Learned:** how to break down complex and difficult concepts and communicate them effectively for better understanding, work within a large chain and team collaboration

Technical Projects:

Sign Collaborative

Jan 2024 – Jan 2024

Technologies: Python, keras, OpenCV, tkinter, pyttsx3, speechRecognition, threading

- Created a real time ASL to text software with TTS and speech-to-text capabilities.
- Hyperparameter tuned and trained a CNN model using keras with several convolution + pooling layers, SoftMax activation for the last layer, multiclass cross-entropy loss function and Adam optimiser to gain ~ 83% accuracy.
- Other models such as Random Forest, SVC and K-NN were trained on images represented as tensors using OpenCV however, these models were over/underfitting even after hyperparameter tuning .
- Used a multithreaded (threading) approach to predict each camera frame asynchronously from Text-to-Speech (pyttsx3) or speech-to-text (speechRecognition) functions.

Word Cloud Generator

Jan 2024 – Apr 2024

Technologies: Python, Django, BeautifulSoup, Sentence-Transformers, Vector DB (Pinecone), AWS (EC2),

- Engineered a full-stack quote recommendation system using Django, implementing web scraping with BeautifulSoup and sentence embedding using paraphrase-MiniLM-L6-v2 model to create 384-dimensional vectors
- Integrated with Pinecone's vector database on GCP for efficient similarity search, and developed custom matplotlib-based visualization stochastic algorithm for dynamic and unique word cloud generation.
- Deployed and maintained the application on AWS EC2 with automated scripts, implementing secure user authentication and CSRF protection

Credit card default classifier

May 2023 – Jun 2023

Technologies: Python, Scikit-learn, Pandas, Matplotlib, NumPy, SHAP, Git, JupyterLab

- Performed Exploratory data analysis and build preprocessing pipelines to transform input data into usable formats for different machine learning models.
- Built ML pipelines for linear (logistic regression) and non-linear (SVM-RBF and Random Forest) classifiers.
- Performed hyperparameter optimisation using grid search and randomised search for different models.
- Forward selection with a logistic regression model is used for feature selection.
- The model yielded an 82% accuracy with low recall due to class imbalance.

Volunteer:

UBC AI Club

Sept 2023 – Apr 2024

University of British Columbia, Vancouver

- Responsible for developing a comprehensive and diverse AI and Machine Learning curriculum.
- Organizing and conducting workshops to educate club members on practical AI/ML tool usage within their respective domains.
- Collaborating with industry professionals and subject matter experts to arrange talks and lectures.
- **Learned:** Curriculum designing, presentation skills, problem solving, collaborative teamwork, project planning and management

ISB Philanthropy Organisation

Sept 2019 - Feb 2020

International School of Berne, Konolfingen

- Responsible for raising money for BM Twin School in Ethiopia and help them to continue their education (Especially during the COVID 19 pandemic peak)
- Organised and facilitated events such as a leg waxing events for with an interactive fundraising element to raise more money for the school.
- Organised Bake sales every Friday through the ski-train to raise more awareness and raising funds.
- **Learned:** Organisational skills, Collaboration between different board members, Marketing skills

Education:

University of British Columbia

Aug 2021 – Present

- 4th year of BSc in Computer Science and Statistics (minor in data science)
- **Achievements:** Faculty of Student International Scholarship, Trek excellence scholarship award
- **Relevant courses:** Software engineering, Intermediate Algorithms, Computer hardware and operating systems, Statistical modelling, Data mining and machine learning, Computer vision, Intelligent systems (reinforcement learning), Time series and forecasting.

Google Data Analytics Certificate

Jul 2022 - Dec 2022