Anshul Kaushal

Mississauga, ON, CA · anshul.kaushal1997@gmail.com · +19058671029 ·
https://anshul-kaushal.github.io/anshul_kaushal/

Education

The University of British Columbia — Vancouver, BC
Master of Data Science Computational Linguistics *GPA: 4.33/4.33* — Sep 2021 - Jun 2022

Humber College — Toronto, ON
Certificate of Achievement Web Design and Development *GPA: 79.3/100* — Jan 2021 - Aug 2021

Punjabi University — Patiala, PB, IN
Bachelor of Technology Computer Science and Engineering *GPA: 8.24/10* — Aug 2015 - Jun 2019

Experience

Katalon Inc. — Vancouver, BC
Data Scientist Intern - Natural Language Processing, Capstone Project — May 2022 - Jun 2022

- To help software testers, implemented an end-end prototype pipeline that translates software test cases written in plain English to a programmable test script written in the Groovy programming language using NLP methodology with only a limited set of examples.
- Leveraged Python data structures and semantically parsed representations of the test cases to navigate unique object paths of the software elements (buttons, text-boxes etc.), with a 100% success-rate (uniquely: 90% and a list of 2-3 possible values: 10%)
- Provided the organisation with examples for writing relatively articulate test case instructions in order to facilitate easier translation.

Infosys Pvt Ltd. — Mysore, KA, IN
System Engineer, Trainee — Jan 2019 - May 2019

- Learnt and applied technologies pertaining to the domain of software engineering including Python, SQL, Java, Java Spring Boot, REST APIs, TypeScript, Angular etc.
- Successfully implemented a web application to assist data analysts at Infosys, working on a product, to monitor the synchronisation of data (from multiple data-points) when transmitted from a local system to a server.
- Recognised as a 'high performer' by the organisation for performing well in the conducted theoretical and practical exams during the training.

Skills

Programming Languages:	Python, R, Java, C++, C
Data Management and Retrieval:	PostgreSQL, MySQL, MongoDB, Elasticsearch
Data Analysis:	Python (Pandas, NumPy), R (tidyverse, lubridate), Excel
Data Visualization:	Tableau, Python (Altair, Seaborn), R (ggplot2)
Machine Learning:	Scikit-learn, XGBoost, R packages
Deep Learning:	PyTorch, Huggingface, SpaCy, Gensim
Neural Net Models:	RNN, LSTM, CNN
Large Language Models:	Transformers (BERT, GPT-2, RoBERTa), T5
Supervised Learning:	Decision Trees, Random Forest, SVM, NN, Naive Byes, Regression, Boosting
Unsupervised Learning:	Clustering (K-means), Dimensionality Reduction (PCA)
Data Engineering:	Beautiful Soup
Software Engineering:	Java Spring Boot, RESTful WEB Services, Postman, Docker, Git (Github)
Web Development:	HTML, CSS, Javascript, Typescript, Angular, React, Node js
Development Environment:	Jupyter Lab, Google Colab, R-Studio

Projects

Line and Length Feature Extractor: Cricket Commentary *PyTorch, Huggingface, Pandas, BERT*
https://github.com/anshul-kaushal/line_length_feature_extraction_cricket_commentary
Built a model to extract line and length features of a ball delivery from cricket text commentary with a precision of 93% and an F1-score of 86%, using merely 100 self-annotated examples and a semi-supervised learning strategy.

Analyzing Customer Churn *Tableau*
https://public.tableau.com/app/profile/anshul.kaushal/viz/AnalyzingDatabelsCustomerChurn/Story1

Analyzed a dataset from scratch to create various interactive visualizations, further combined into multiple dashboards, and presented as a story, to help highlight KPIs and answer questions pertaining to the customer churn at Databel.

Bot Tweets Classifier *PyTorch, Huggingface, Pandas, BERT, RoBERTa, T5, GRU, M2M100*
https://github.com/anshul-kaushal/bot_tweets_classifier
Trained numerous deep learning models to detect artificially generated English and French tweets, and achieved a 90% F1-score for tweets generated by non-GPT-2 bots.

Movie Script Generator *PyTorch, Huggingface, Pandas, Beautiful Soup, GPT-2*
https://github.com/anshul-kaushal/movie_screenplay_generator
Built a movie script generator that can generate a movie screenplay with apt format, given a prompt, using the GPT-2 model fine-tuned on horror movie scripts scraped from the internet.