SAKET THAVANANI

University of Toronto, Canada

saket.thavanani@mail.utoronto.ca (+1) 4372145011
/saket-thavanani-b1a149a0/

EXPERIENCE

Data Scientist: Internship

Fidelity Canada:

June 2020 – December 2020 Toronto

- Developed a personalized Recommender system for the Fidelity Sales team using Collaborative filtering techniques to complement the institution's currently existing Market Basket model.
- Built the firm's next generation customer segmentation model using Agglomerative Clustering and Gaussian Mixture Model.
- Designed an Ensemble approach for developing Recommendation engine using three different machine learning algorithms- ALS, Neural Collaborative Filtering and Clustering.
- Designed an evaluation methodology using AUC score, Intra-list similarity and masking to validate the performance of recommender system.

Graduate Teaching Assistant, University of Toronto

Course: Data Analytics and Machine Learning

January 2020 – Present Toronto

- Responsible for conducting the weekly lab sessions, assisting students with their code.
- Actively monitored course forum on piazza and helped students with the core concepts of data science and Machine Learning algorithms.
- In charge for designing course projects and hosting them on Github. Other duties include grading and maintaining course content on the school's course website.

PROJECTS

Predicting Popularity of Reddit Posts

Academic Project

June 2020 - September 2020 Toronto

- Developed an Artificial Intelligence Algorithm that can Predict popularity of Reddit Posts leveraging the Big Data concepts in Pyspark/Scala using Spark ML, Spark NLP as predominant Machine learning packages.
- Extensive Feature Engineering was performed on the dataset to extract innovative features like - Webscrapped Comments from URL of the Post, Window features, Glove Embeddings, Word Embeddings, BERT Embeddings, etc.
- Implemented two different machine learning models- Random Forest and Gradient Boosting, With Gradient Boosting outperforming the former one with a Mean Absolute Error of 8.01 on the Test set.

Fact Checking

Data Cup Competition

October 2019 - December 2019 Toronto

- Developed an Artificial Intelligence algorithm that can rate a claim as true, partly true and false.
- Three models were implemented - LSTM, Passive Aggressive Algorithm and BERT. With BERT model achieving the highest accuracy of 0.60 on test set.

EDUCATION

Master of Engineering : Emphasis in Machine Learning and Data Science

University of Toronto

Jan 2019 – December2020 GPA: 4.00/4.00

B.Tech + M.Tech : Metallurgical and Materials Engineering

Indian Institute of Technology,Kharagpur

July 2013 - May 2018 CGPA: 9.26/10.00

ACHIEVEMENTS

- **Data Cup**
  Stood among Top 30 teams in Canada in Data Cup Competition
- **Charpak Scholarship**
  Selected among the top 25 students from various branches of engineering all over India for the Charpak Research Internship Program
- **IIT JEE Advanced**
  Stood among top 0.02 Percentile out of approximately 1.5 million students.
- **JEE Mains**
  Secured an all India Rank of 520 out of approximately 1.5 million student
- **GATE Exam**
  Secured an all India Rank of 32 out of approximately 50k students.

SKILLS

Python SQL R PySpark Scala
Flask HTML Pytorch Keras Scikit-Learn
NLTK MLlib Spark ML Spark Nlp
Matplotlib Seaborn Plotly ggplot2
Tableau SciPy NumPy pandas
Databricks Hadoop Hive Git

STRENGTHS

Hard-working Team Player Self-Motivated
Problem Solving Good Communication
Time Management Creative Thinking
Strong Ability to work under Pressure

PROJECTS

Natural Language Processing for Social Media Analytics

Academic Project, Course: Data Science and Analytics

September 2019 - Dec 2019 Toronto

- This project predicts the sentiments of tweets based upon various political affiliations.
- A simple procedure was designed that determines the political party (Liberal, Conservative, NDP, None) of a given tweet and applied this procedure to all the tweets in the 2019 Canadian elections dataset.
- The features were extracted using using three different techniques (Bag of words, N-Grams and TF-IDF)
- Five different models were implemented - Support Vector Machines, Logistic Regression, Naive Bayes, Random Forest and Neural Networks, with Neural networks achieving the highest accuracy of 0.63 on test set.

Deep Learning to Predict Polymers Glass Temperature

Academic Project, Course: AI in Materials Design

May 2020 - July 2020 Toronto

- This project used image of the given polymeric compound to predict its glass transition temperature using the SMILES encoding of chemical compounds.
- Two models were implemented -Convolution Neural Networks and Fully Connected Neural Networks, With Full Fully Connected Neural Networks achieving lowest Mean Percentage Error of 5%.

Image Classification

Academic Project

January 2020 - April 2020 Toronto

- This projects aims to compare performance of image classification task using Convolutional Neural Networks with 8 sets of optimization algorithms GD, SGD, Momentum, Adam, Adagrad, Adadelta, RMSprop, Nestrov Momentum.
- Each Algorithm was implemented and evaluated mathematically using Numpy. The idea was to come up with a new formulation in existing algorithms which can yield best results. Another aim was to arrive at the best architecture of CNN.

Salary Classification of Data Scientists

Kaggle Competition

September 2019 - Dec 2019 Toronto

- The project aims to train, validate and tune the multi-class classification problem that can classify, given a set of responses by a data scientist, what a survey respondent's currently yearly compensation bucket is.
- Performed comprehensive data cleaning of a large dataset with multiple data types,and feature engineering. EDA was performed using Seaborn, Matplotlib to understand the correlation between feature and the target variable
- Developed a multi-class classification model using Logistic Regression algorithm yielding an accuracy of 0.51 on test set.

COURSEWORK

- Design and Analysis of Algorithms
- Data Science and Analytics
- Data Analytics and Machine Learning
- Big Data Science
- Introduction to Machine Learning
- Convex Optimization
- Programming in C and Data Structures
- Probability and Stochastic Processes
- Operating Systems

CERTIFICATIONS

- Big Data with Pyspark
- Data Structures and Algorithms
- Natural Language Processing
- SQL for Data Science
- R for Data Science
- Introduction to Tableau

PROFILE LINKS

•Linkedin
/saket-thavanani-b1a149a0/

•Github
Saket-Uoft

•Medium
https://medium.com/@saketuoft

•Towards Data Science
https://towardsdatascience.com/@saketuoft

•LeetCode
https://leetcode.com/saketthavanani/