PRANSHU KUMAR
DATA SCIENTIST

Employment

Northeastern University Experiential Network (XN) — Boston, MA
Data Analyst — Jan. 2020 to Mar. 2020
- collaborated for a short-term XN project for contact sourcing for a Private Equity firm, the Allston Group, Allston, MA.
- web scraped data, acquired physical therapy practice details in the Northeast through roll-up strategy.
- compiled spreadsheets with ABA and pediatric therapy companies details including contact information and geographies in all US states.
- aligned with the firm's strategy, analyzed the private equity industry and dealt with processes for client acquisition.

Intel — Dehradun, India
Summer Trainee — July 2018 to Sept. 2018
- created complex machine learning models for successful completion of 'AI 101' Intel Certified course with a 95 percentile score.
- covered Neural Network architectures, convolutional networks, and recurrent networks to complete Intel certified Deep Learning training with 96% assessment grade.
- Created a machine learning project to perform statistical analysis, predict FIFA 2018's Best XI Players.

Projects

Personalized Cancer Diagnosis
- multi-class -classified given genetic variations/mutations based on evidence from text-based clinical literature.
- achieved log loss values of 1.15 & 1.03 for Naïve Bayes and K-Nearest Neighbors as baseline models
- trained Logistic Regression with Count Vectorizer features, unigrams and bi-grams, Linear SVM, achieved average 1.06 log-loss
- trained Random Forest with one-hot encoding for hyperparameter tuning
- achieved 0.53 and 0.83 log-loss for training Stacking & Maximum Voting Classifiers

NYC Taxi Demand Prediction
- predicted cabs pickup demand in 10 minutes time frame, given region co-ordinates.
- applied K-Means clustering using GridSearch that found minimum inter-cluster distance for given NYC region
- analyzed top amplitudes & corresponding frequencies using the time-series Fourier transform plot
- used Weighted, Exponential Moving Averages as baseline models
- used Linear Regression, RandomForest, and XGBoost with Grid, Random Search, achieved 12% MAPE for both train & test data

Facebook Friend Recommendation using Graph Mining
- supervised machine learning problem that predicted missing links from the given directed social graph.
- generated training samples for good & bad links using page rank, Katz, score, Adar index directed graph techniques.
- achieved an F-1 score of 0.9241 for Random Forest, 0.9327 for Gradient Boosted Decision Tree for predicting links.

Amazon Fashion Discovery Engine
- content-based recommendation engine for women's apparel on Amazon using text, image data scraped from Amazon product advertising API.
- encoded text-based features using Bag of Words(BoW), Tf-idf & idf techniques.
- made semantic-based predictions using Average Word2vec, and idf Weight.Word2vec techniques.
- predicted a wide range of apparels using CNN on feature extracted image vector data.

Awards

IBM · IBM Data Science Professional Certificate — Jan. 2020
- developed and honed Data Science & Machine Learning skills.
- completed cloud-based labs and assignments including a Capstone project for skills demonstration.

Northeastern University · Experiential Network badge — May 2020
- successfully completed a sponsored professional project for a real-world organization.
- presented complete project deliverable sponsor solution with satisfied scoped business needs.

Coursera · Deep Learning Essentials with Keras — Jan. 2020
- demonstrated understanding of Supervised, Unsupervised deep learning models including autoencoders, restricted Boltzmann machines.
- successfully built deep learning models, networks using Keras library.

IBM · Machine Learning with Apache Spark — Jan. 2020
- successfully solved data science and machine learning problems involving Big Data using Apache Spark.

Contact

pranshu1921@gmail.com
pranshu1921.github.io
Boston, MA, USA
linkedin.com/in/pranshu-kumar
pranshu1921

Education

Northeastern University — Sept. 2019 to Current
MS Analytics
Relevant Courses : Intermediate Analytics, Data Mining Applications, Predictive Analytics

University of Petroleum and Energy Studies, Dehradun, India — July 2015 to July 2019
BS Computer Science
Relevant Courses: Artificial Intelligence, Advanced Database Management Systems

Skills

PROGRAMMING
Python (pandas, numpy, scikit-learn, seaborn, nltk etc.)
SQL
R
Java

UNSUPERVISED LEARNING
K-Means Clustering
DBSCAN
Recommendation Systems

SUPERVISED MACHINE LEARNING
Decision Trees
Naive Bayes
KNN
Linear/Logistic Regression
Linear/Logistic Regression
KNN
Naive Bayes
Decision Trees

STATISTICS
Decsriptive Statistics
Exploratory Data Analysis
Inferential Statistics

DATA VISUALIZATION
Matplotlib
Tableau
Power BI

DEEP LEARNING
Natural Language Processing
Convolutional Neural Networks
Tensorflow
Keras
Autoencoders