Bhavika Tekwani

btekwani@gmu.edu | Phone: 571-326-4698
github.com/bhavika| linkedin.com/in/bhavikatekwani

EDUCATION

George Mason University, Virginia August 2016 - May 2018
Master of Science, Computer Science
Relevant coursework: Theory & Applications of Data Mining, Data Mining on Multimedia Data, Pattern Recognition, Applied Statistics, Mining Massive Datasets, Distributed Software Engineering, Analysis of Algorithms

University of Mumbai, India May 2010 - June 2014
Bachelor of Engineering, Computer Engineering

TECHNICAL SKILLS

Languages: Python, Java, SQL, C#, C++, Matlab, R
Tools: Tableau, IPython, Weka, Docker, AWS, Git, Apache Airflow, CI/CD (Travis, CircleCI)
Techniques: Clustering, classification, natural language processing, regression, matrix factorization
Libraries: scikit-learn, Numpy, Pandas, Scipy, GraphLab, matplotlib, seaborn, gensim
Beginner level knowledge of Tensorflow & Keras.

WORK EXPERIENCE

Software Engineer Intern- Data, Udacity *Mountain View, CA* June 2017 - August 2017
- Improved organizational access to data by building a Slack bot with Flask and SQLAlchemy.
- Built an ETL system to migrate platform events to Amplitude for real-time analytics.
- Analyzed 300,000+ events to gain insights for Udacity products - Mentorship & Projects.
- Built data pipelines using Apache Airflow (Python).

GIS Programmer, George Mason University August 2016 - present
- Developed George Mason University's OpenGeoportal using Java, Apache Solr, Python.
- Lead workshops on ArcGIS, QGIS, Python and CartoDB to train students and faculty.
- Assisted students with framing research questions and designing projects based around GIS and data analysis

Senior Software Engineer, Capgemini June 2015 - July 2016
- Improved the performance of SQL stored procedures by 0.5-2x.
- Developed features for forecasting, asset management and resource allocation.
- Improved perceived page load performance on dashboards by 30%.
- Reduced page load speed by 6s for frequently accessed pages using caching mechanisms.

Software Engineer, Capgemini June 2014 - June 2015
- Developed a Java/J2EE application for a Fortune 500 engineering client.
- Built a prototype for a product lifecycle management application using Java.

PROJECTS

Raven: An open-source CLI tool to manage your Spotify music library.
URL: https://github.com/bhavika/raven

scuzzy: A Scala library for dead simple fuzzy text matching. Currently a work in progress.
URL: https://github.com/bhavika/scuzzy

Quora Question Pairs: A Kaggle competition about detecting duplicate questions on Quora. Achieved a log loss of 0.324 on the leaderboard using an LSTM built with Keras and TensorFlow.
URL: https://github.com/bhavika/Quora_QuestionPairs

Music Mood Classification Using the Million Song Dataset: Used audio features and machine learning for to classify 7000+ popular songs by mood with ~75% accuracy. Techniques: segment aggregation, XGBoost, Random Forest, Support Vector Machines
URL: https://github.com/bhavika/JoyDivision

Dynamic Time Warping (DTW): Implemented the DTW algorithm with lower bounding (Keogh's technique) to provide a distance metric for time series data.
URL: https://github.com/bhavika/DMMultimedia/tree/master/HW1

Motif Discovery & Classification of Time Series Data: Adapted the bag of words model as a bag-of-patterns to classify time series data where features were the SAX representations of discovered motifs to compare accuracy against several other benchmarks.
URL: https://github.com/bhavika/DMMultimedia/tree/master/HW2

Movie Recommender System: Predicted movie ratings given by a user through item based collaborative filtering. Achieved an RMSE of 0.76. Techniques: factorization machines
URL: https://github.com/bhavika/DataMiningProjects/tree/master/Recommender

Amazon Review Classification: Implemented the k-Nearest Neighbour algorithm to classify Amazon reviews by sentiment. Techniques: TF-IDF, Singular Value Decomposition (SVD)
URL: https://github.com/bhavika/DataMiningProjects/tree/master/AMZReview

Drug Activity Prediction: Achieved an F1 score of 0.73 while classifying drugs as active/inactive in an imbalanced dataset. Techniques: Stochastic Gradient Descent (SGD)
URL: https://github.com/bhavika/DataMiningProjects/tree/master/DrugActivity

scipy: I've made a few contributions to the project that are shown as closed PRs on the URL below.
URL: http://bit.do/scipy-prs