Nikhil Takappa Saunshi

669-204-5721 | 201 S 4th Street, San José, CA - 95112
LinkedIn | GitHub | nikhilts19@gmail.com

EDUCATION

Master of Science in Computer Engineering - San José State University, San José, CA, USA Dec 2019
Courses: Data Structures and Algorithms in C++, Machine Learning, Data Mining, Large-Scale Analytics

Bachelor of Engineering in Electronics and Communication - M.S. Ramaiah Institute of Technology, Bangalore, India Aug 2015
Courses: OOP and Data Structures, Operating Systems, Fundamentals of Computing, Neural Network and Fuzzy Systems

TECHNICAL SKILLS

- **Strong Background:** Machine Learning, Classification, Regression, Data Mining, Wrangling & Visualization
- **Programming Languages:** Python, SQL, C++, Data Structures & Algorithms, Javascript
- **Natural Language Processing:** Gensim, NLTK, Beautiful Soup, Topic Modeling (LDA)
- **Frameworks & APIs:** Pandas, NumPy, SciPy, Keras, Scikit-learn, Matplotlib, Seaborn
- **Tools:** Tableau, Jupyter Notebook, Anaconda suite, Eclipse, AWS, LaTeX, PyCharm, Git, GitHub, SFDC, RxClaim, EzQuery, EzTest
- **Working level proficiency:** Deep Learning, Machine Learning Modeling, Statistical Data Analysis, Clustering, CNN

WORK EXPERIENCE

Research Assistant - Machine Learning, Tower Foundation, San José State University, USA March 2019 - Present

- **Data Collection (SERP Mining):** Developed a web scraping script in python to scrape top 100 ranked search result from Google search engine. Designed the application to generate labels automatically with the Titles and Descriptions, establishing a curated data set for further data analysis.
- **Natural Language Processing:** Analyzed the data with exploratory data analysis (EDA) and applied Natural Language Processing (NLP) techniques using NLTK library for text pre-processing.
- **Feature Engineering:** Extensively researched and worked on feature engineering to implement TF-IDF vectorizer and topic modeling using Latent Dirichlet Allocation. Converted topics to feature vectors by generating sparse matrix for ML modeling.
- **Classification:** Applied Cross-Validation and Grid Search to find optimized hyper-parameters for the model. Built end-to-end Machine Learning pipeline to achieve multi-class text classification with a weighted average f1-score of 91%.
- **Topic Modeling:** Currently working on evaluation of the LDA model to improve the classification accuracy of the ML model.

Programmer Analyst, Cognizant Technology Solutions, India May 2016 - June 2017

- Interpreted and reviewed the accuracy of payments involved in batch processing of claims using RxClaim system, EzQuery and EzTest. Analyzed the data and presented it graphically using charts in MS Excel to make meaningful decisions.
- Tracked and managed large volumes of customer, product, marketing and sales data using salesforce (SFDC).

ACADEMIC PROJECTS

Sentiment Analysis and Visualization using Yelp Data [Python, Pandas, NLTK, Seaborn, Scikit-learn, Classification]

- Performed Exploratory Data Analysis on Business and Review datasets to understand the data and relationship between the two datasets. Merged the two datasets using pandas dataframe and labeled the data for Machine Learning modeling.
- Applied various classification techniques to analyze sentiment on restaurant reviews to identify the negative/positive aspects of the restaurant. Compared the ranking against the Yelp's 5-stars ranking.

Song Recommendation System [Python, Pandas, Numpy, Filtering, Similarity Metrics, Scikit-learn, Matplotlib]

- Researched and developed a hybrid song recommendation system based on collaborative filtering and content-based filtering to improve user engagement. Evaluated the model using using evaluation metrices like precision and recall.

Find My Keys Hands-Free [AWS Lambda, Python, Raspberry Pi]

- Designing and developing a voice-activated personal assistant involving an AVS device. Built the AVS client using Raspberry-Pi.
- Created a custom skill to invoke alerts on the bluetooth tag using python. Deployed the skill on AWS lambda to locate the keys.

Digit Recognizer [Python, Keras, Tensorflow, PCA, CNN, Pandas, Numpy, Matplotlib, TSNE, Random Forest]

- Leveraged various Machine Learning models to implement digit recognizer/classifier using Scikit-learn & Keras with Tensorflow backend in python for accurate detection of handwritten digits using MNIST dataset.
- Implemented the above algorithms to identify the best working model based on optimal parameters, runtime and test error. Evaluated the model on Kaggle and achieved an accuracy of 99% to stand in the top 30% of the competition.

PUBLICATION

- 2019 IEEE International Conference on Big Data - A SERP Mining Approach for Topic Classification of DNS Requests.

AWARD

Runner-up of Cisco Hackathon (enhackathon 2019) - The Bridge to Innovation [Javascript, Cisco DNAC] July 2019

- Developed customizable applets using Javascript to create alerts in the Cisco DNAC. Utilized telemetry data to create alerts.