ANKOOR BHAGAT

https://github.com/ankoorb
http://ankoorb.blogspot.com
http://www.linkedin.com/in/ankoorb

☎ +1 (949) 331-2374
✉ ankooris@gmail.com
Los Angeles, CA-90049

SUMMARY

A talented, passionate and self motivated data scientist. M.S, Ph.D. in Engineering. Proficient in using Python, R and MATLAB

SKILLS

Programming: Python (Pandas, NumPy, sciKit-learn, IPython), Flask, SQL, Spark (PySpark), MATLAB, R, D3.js (Intermediate) and familiar with JavaScript, HTML and CSS

Applied Machine Learning, Regression, Bayesian Data Analysis (PyMC), Natural Language Processing (NLTK & Gensim), Deep Learning (TensorFlow), Interactive Data Visualization, Dimensionality Reduction, A/B Testing, Clustering Algorithms, Optimization, Transportation Algorithms, Engineering Economics, Transportation Systems Analysis, Traffic Simulation, Emission and Air Dispersion Modeling, LaTeX

SIDE PROJECTS

- Developed and deployed an Image Classification web app that uses 3-layer Convolutional Neural Network to classify whether an uploaded image contains either a dog or a cat: Click here to try or visit http://dogvscat.pythonanywhere.com/
- Developed and deployed a Movie Recommendation web app using Item-Item Collaborative Filtering and Alternating Least Squares Algorithm: Click here to try or visit http://amovierecommender.pythonanywhere.com/
- Developed and deployed a House Value Prediction web app using Ridge Regression: Click here to try or visit http://ahouse-value.pythonanywhere.com/
- Interactive data visualization side projects using D3: Click here to check or visit https://ankoorb.github.io/

EXPERIENCE

Data Scientist, mPulse Mobile, Encino, (Mar 2016 - October 2016)

- Performed exploratory data analysis (PostgreSQL) and created data visualizations to understand consumer behavior
- Developed Support Vector Machine models (using scikit-learn) to classify text messages. Implemented models in API using Python and Flask for text message solution workflows (Demo Web App link)
- Analyzed text message data and evaluated the performance of Logistic Regression, Naive Bayes, Decision Trees, Random Forest and Support Vector Machine models (using Scikit-Learn) for text message classification
- Evaluated k-means clustering, Hierarchical Clustering, Non Negative Matrix Factorization, and Topic Modeling (gensim) for text message labeling. Recommended using Amazon Mechanical Turk to label text messages for training classification models
- Evaluated the performance of Natural Language Processing API's (Google Prediction, IBM Watson, and a few others) for sentiment analysis and text classification
- Developed two regular expression API's with Python and Flask to parse human readable date/time for text message solution workflows
- Developed an API with Python and Flask to report current Air Quality Index by ZIP code for text message solution workflow
- Coded JavaScript (D3.js) scripts to create interactive data visualizations for dashboards (Demo Web App link)

Intern, Sarakki Associates Inc., Santa Ana, (Sept 2014 - Mar 2016)

- Coded Python scripts to estimate probability distribution of toll revenue forecast to understand risk and uncertainty in toll road projects (Contract from a potential start-up)
- Determined the revenue generating potential of Real-time Traffic Archival Data Management System project

Data Science Fellow, NewMet Data Science Bootcamp, Los Angeles, (Sept 2015 - Nov 2015) (GitHub repository link)

- Coded Python scripts to scrape and integrate data from various sources, including trash diversion rate, census data, political affiliation, crime, solar energy data
- Performed Exploratory Data Analysis (EDA) and Feature Engineering by developing various indices to reduce over 2000 features to 45 features
- Applied various machine learning algorithms including PCA, k-means, Decision Tree, Random Forest to explain factors affecting waste management and made a recommendation list of the cities for Earth Advocacy Project Nonprofit Organization to target

Graduate Student Researcher, University of California, Irvine, (Jan 2009 - June 2014)

PROJECTS

- **Network Augmentation Algorithm**
  - Designed, implemented and tested network augmentation algorithm to reduce Origin Destination (OD) matrix estimation time (from over 100 iterations to less than 30) using MATLAB

- **Freeway Accident Data Analysis**
  - Coded Matlab and R scripts to analyze 3 years of Los Angeles freeway accident data and estimate the temporal risk of accidents on I-710 and I-110 freeways. Used Python to implement a hierarchical bayesian model to detect change in accident rates
- **Environmental and Health Impacts of PierPASS Program**
  - Analyzed 60+ GB of traffic simulation trajectory data to model vehicular emissions and estimate spatio-temporal impacts of air pollution from freight deliveries using MATLAB

Teaching Assistant, University of California, Irvine, (Jan 2009 - June 2014)

- Instructed undergraduate students in Economics, Statistics, Linear Regression, Linear Programming and Non-linear Optimization courses

EDUCATION

Doctor of Philosophy in Transportation Systems Engineering — June 2014
University of California, Irvine — GPA 3.8

Master of Science in Transportation Systems Engineering — December 2007
University of California, Irvine — GPA 3.6

Bachelor of Engineering in Civil Engineering — June 2003
Nagpur University, India — First Division

MISCELLANEOUS

- Peer reviewed papers in transportation research (Complete list available upon request)
- Four time semi-finalist in The Data Incubator's Data Science Fellowship program
- Delivered numerous lectures to students, conference presentations to researchers and industry professionals
- Supervised numerous graduate students' M.S. theses and mentored several successful undergraduate students conducting research
- 16^{th} Annual UCTC Student Conference Poster Committee Chair, UC Irvine, 2010