1236, East Lemon Street,
Tempe, Arizona

RAMAN AHUJA

rvahuja@asu.edu
Linkedin
Personal Website

EDUCATION

Arizona State University **MS in Computer Science**

- **Coursework**: Foundation of Algorithms, Fundamentals of Statistical learning, Natural Language Processing

Visvesvaraya National Institute of Technology **B.Tech in Computer Science** **C.G.P.A : 7.80/10.0**

- **Electives:** Artificial Intelligence, Neuro Fuzzy Techniques, Information Retrieval, Data Mining

WORK EXPERIENCE

Numerify India, Bangalore **Associate Software Developer** **June 2016-June 2018**

Performance Optimizing Engine for Amazon Redshift

- Developed an intelligent distribution and sorting mechanism for data on clusters by leveraging Master-slave architecture
- Implemented compression encoding to reduce storage and disk I/O, increased real time analytics performance by 60%
- Developed a framework to reclaim fragmented disk space across VPC's and achieved 40% gain in ETL job processing

Feedback Driven Decision Making

- Developed an extractor pipeline for Microsoft-Yammer. Used NLP for parsing feedbacks
- Trained ML models for opinion analysis of feedbacks. Built intuitive visual dashboards by developing Micro strategy cubes

Healthcare Infrastructure Outage Analyzer

- Built a predictive model which progressively gets trained on historic incidents, to forecast the critical client outages
- Architectured the datawarehouse model and developed an ETL pipeline for the outage analysis and preventive measures
- Developed interactive dashboards to find actionable insights on outage trends which elevated incident resolution rate

Centre of Excellence – VNIT **Research Intern** **May 2014 – Jul 2014**

- Developed a handled device to make eyes learn and discriminate colors for curing partial color blindness

TECHNICAL PROJECTS

Increasing the efficiency of Naïve Bayes model for Sentiment Analysis

- Factored in feature extraction techniques, Negation Handling and Word N Grams, trained model on IMDB data
- To handle new reviews, an adaptive approach was devised using Sentiwordnet based self-learning algorithm
- Obtained accuracy of 85.1% for linear complexity model, at par, to the support vectors machines and neural models

Criminal Hotspots – Data Mining

- Developed an application for interactive analysis of historic crimes by hotspot detection and classification of risk terrains using clustering algorithms in R based on geographical information and various criminal attributes

Fake Mail Detection

- Developed spam email classifier factored in preprocessing, feature extraction and trained on SVM to test accuracy 98.5%

Indexing Tool – Information Retrieval

- Developed a distributed system using Hadoop MapReduce framework for construction of inverted index on large-scale documents to improve search query performance and also relevancy by pre-processing using NLP constructs

KAGGLE PROJECTS

Intruder Detection Through Webpage Session Tracking *[Python(numpy, pandas, scipy, scikit-learn, matplotlib)]*

- Predicted whether a webpage session belongs to legitimate user or someone else, by training ML model on sequence of webpages attended consequently by the person in the history, used sequential pattern mining techniques

Flight Delays *[Python (scikit-learn, pandas, numpy)]*

- Predicted whether a flight will be delayed by more than 15 minutes, improved model performance by onehotencoding

House Price Prediction *[Python (numpy, pandas, scikit-learn – RandomForestRegressor, matplotlib)]* (Kaggle Kernel)

- Predicted sales price of each home, used feature engineering and advanced regression techniques to reach RMSE 0.12

SKILLS

- Java, Python, C, Spark, numpy, pandas, scikit-learn, AWS(S3,REDSHIFT,EC2,EMR,RDS), Mysql, Matlab Linux, Git, Jenkins

TECHNICAL TRAINING EXPERIENCE

- Data Structures and Algorithms - University California, San Diego - Daniel M Kane
- Machine Learning – Stanford University – Andrew Ng