Ramesh Oswal

ramesho@andrew.cmu.edu **linkedin.com/in/ramesh-oswal** **rameshoswal.tumblr.com** **Phone - (412) 587-3598**

Objective – Seeking roles in the field of Data Science specializing in Machine Learning and Big Data Analytics

Education

- **Master of Science – LTI's BIC (School of Computer Science) Major – Machine Learning** **Aug'16 - May'18 (Expected)**
  Carnegie Mellon University, PA
- **Bachelors in Computer Engineering** **Aug'10 - May'14**
  University of Pune,IN.

Relevant Coursework **(*In Progress)**

- Math for Machine Learning
- Intro to Machine Learning
- Machine Learning for Signal Processing*
- Machine Learning for Large Datasets*
- Advanced Databases
- Big Data in Practice

Technical Skills

	Proficient	**Familiar**
Core Languages:	Python, Core Java, Visual Basic,	C, C++, R, Javascript, SAS, MATLAB, Scala
Databases:	Hive, Cassandra, Oracle 9i and 10g, MySQL	MS Access, MongoDB
Dev./Productivity Tools:	Turbo C, Informatica, Excel, Anaconda	Jupyter Notebook, Hadoop, Spark, MS VS 2010 & 14

Academic Projects

Data Science Intern **Walmart Labs** **June'17 – Aug'17**

- Working on the Walmart Performance Ads team to optimize the current model used by Walmart to display relevant ads.
- Predicting Click through Rate(CTR) of ads using contextual information resulting in increase in the revenue.
- Feature Engineering (e.g Binning, polynomial and logarithmic feature transformation), identifying new features & performing experiments to tune hyper-parameters.
- **Tools Used – Python, Spark(MLlib), Scala, Hive, Cassandra, Weka**

Movie Recommendation System using MovieLens Dataset (Carnegie Mellon University) **May'17 - June'17**

- Used the Matric Factorization Technique to recommend movies to users following the Netflix Prize Winner's Strategy on the Movie Lens Dataset consisting of 1 million ratings as training set.
- Implemented the Alternating Least Squares Optimizing Technique to solve the "RMSE" Objective Function.
- Performed Experimental Analysis to tune hyperparamaters like K, lambda etc.
- **Tool Used: Spyder, Python (NumPy, matplotlib, SciPy)**

Home Depot Product Search Relevance (Carnegie Mellon University) **May'17 - June'17**

- Performed feature engineering like cosine similarity, edit distance etc. using NLP techniques like word embeddings on the unstructured dataset consisting of Product Description and Attributes.
- Data preprocessing like Stop word removal, Stemming, and typo correction were performed before feature engineering.
- Used Machine Learning Algos like RandomForest Regressor and Linear Regression to score each search query.
- **Tool Used: Python (NumPy, matplotlib), Big Data/Distributed sytems -Spark – Pyspark, MongoDB**

Super Fridge: Automated Grocery List using Object Detection in Refrigerator(CMU) **Mar'17-Apr'17**

- Built an application running on Raspberry Pi using the camera module to detect objects in a Refrigerator and creating a Grocery List for missing items.
- Built modules consuming Clarifai Api used for object detection using a picture clicked from Pi camera and push the grocery list to google drive for users.
- **Tools Used: Python, Raspberry Pi & Camera, Calrifai API (Object Detection), Google Drive API**

Musicon: Music playing based on User Activity Recognition: SteelHacks'17 **24hr – Hackathon (Feb'17)**

- Built an Android app which used Google's Accelerator(Motion Sensor) data to determine User Activity(Brisk Walk, Jogging, Sprint, Standing etc).
- Integrated the User activity recognition module with Spotify API, which played song based on user activity and switched between them.
- **Tools Used: Android JDK, Java, Google Accelerator (motion sensor) API, Spotify API**

ramesho@andrew.cmu.edu

Project Intern **Talencea Inc, Pittsburgh** **Oct'16 – May'17**

- Worked with a Pittsburgh based startup founded by LTI Director Dr. Jaime Carbonell.
- Working on Big Data from different external sources like client and social media platforms and building Skill Repository.
- Data Munging activities include Data Cleanup, Indexing, Classification, Redundancy Removal & etc.

Image classification to classify proteins into subcellular localization patterns (CMU) **Aug'16 - Dec'16**

- Built an Active Learning Framework containing Pool Based Data Access Model, Uncertainty based Querying Strategy and different base learners like SVM, Gaussian NB, KNN and Logistic Regression
- Accuracy score of 0.97 was achieved on test data using SVM as base learner.
- **Tool Used: Spyder, Python (sklearn, NumPy, matplotlib, SciPy)**

Using Probabilistic Graphical Model to forecast Stock Prices for Time Series data (CMU) **Aug'16 - Nov'16**

- Transformed into stationary time series by using log space for removing unequal variances & difference to handle trends.
- Created precision matrix using transformed features and marginalized Precision Matrix for missing data.
- Conclusively was able to predict with minimal error rate the stock prices for Apple by using only 3 days worth of data and stock prices for companies MS, Hecla, NEM.
- **Tool Used: Spyder, Python (NumPy, SciPy)**

Linear and Forward Stage-wise Regression on unknown Dataset (Carnegie Mellon University) **Aug'16 - Nov'16**

- Implemented Linear and Forward Stage-wise Regression from scratch.
- Implemented Feature transformation like One Hot Encoding and Polynomial Feature transformation.
- Cross-validated different train and test dataset combination to get optimized weights for each feature.
- **Tools Used: Jupyter, Python (NumPy, SciPy)**

Load Balancer for OpenFlow compliant SDN architecture (Sponsored by GS Labs Pvt. Ltd) **July '13 - Jun '14**

- Aimed at enhancing s/w load-balancer in distributing traffic based on server capacity by adding generic flows.
- R. Oswal et. al. "A Survey of Past, Present and Future of Software Defined Networking"
- **Tools Used - Mininet with POX controller, OpenVswitch and OpenFlow protocol**

Paper Presentation "A Cloud Framework for Parameter Sweeping Data Mining Application" **Jan '13 – Feb '13**

- Explained the system framework i.e. its architecture and execution mechanism of how parameter sweeping could be achieved in data mining application. Finally, concluded by showing a performance evaluation w.r.t clustering & classification

Professional Experience

Business Operation Associate **ZS Associates Inc.** **Sept'14 – June'16**

- **Master Data Management - Role (Data Steward)**
  - Automated processes like loading client data and QCing client deliverable and performed Ad-hoc analysis.
  - Automation of Processes to reduce response time for file processing by over 80%
  - **Technologies Used – Python, MS Excel, VBA, Informatica Siperian, PL/SQL**
- **Smart Data Quality Management – (Quest '15)** **24hr – Hackathon (Oct'15)**
  - Participated & won in Quest'15 organized by ZS Associates which had 44 participating teams.
  - Designed Product Architecture detailing flow and control of Data pipeline.
  - Implemented an algorithm for "Thomson Tau Method of Outlier detection" to detect outliers
  - **Technologies Used – R, MS Excel, VBA and MS Access**

Summer Intern **Softkoash Solutions Pvt. Ltd** **May '12 – July '12**

- Implemented Microsoft's NerdDinner project as a POC
- Fixed bugs and made changes to proprietary ERP Solution used by customers in production.
- **Technologies Used – C#, Microsoft's .NET Framework, HTML, CSS and JavaScript**

Co-curricular Activities

- Won in Quest '15 (Hackathon at ZS Associates)
- 2nd Best Project - PICT's "Impetus & Conceptus'14"
- 2nd Prize in College TechFest Event 'Network Raptors'
- Best Project in Ops Excellence ZS Associate