Mitali Ingole

312-956-5186 | mitali1210@gmail.com | Chicago, IL

SUMMARY

Data scientist with strong qualitative and quantitative research background interested in applying data science solutions to business challenges; experienced with statistical modeling, relational and NoSQL database programming, and data visualization using R, Python, Java and SQL.

EDUCATION

MASTER'S | DEPAUL UNIVERSITY | JUNE 2017 — **MAJOR**: Computer Science | **MINOR**: Data Science

BACHELOR'S | PUNE UNIVERISTY | JUNE 2011 — **MAJOR**: Computer Science

SKILLS

- **Programming:** C, C++, R, Python, SQL, Java, Scala
- **Software and system:** SAS, SPSS, HDFS, WEKA, Tableau, Alteryx, Mahout, Apache Spark, Microsoft Office Suite
- **Database:** SQL Server, MySQL, SQLite, Oracle SQL, NoSQL, Hive, Pig
- **Modeling:** Regression, Statistical Significance Testing, Machine Learning, Time series, Regularization

EXPERIENCE

MACHINE LEARNING RESEARCHER | DEPAUL UNIVERSITY | 2015 – PRESENT

- Currently conducting machine learning research in statistical analysis, supervised, semi supervised, and unsupervised data mining techniques under the direction of DePaul's Provost Dr. Helmut Epp.
- Leading various lectures and trainings on R, Python, Hadoop, Scala, & Java to university students and company employees.
- Designing course structure for RMaster program using Rmarkdown and knitr focusing on supervised & unsupervised learning, natural language processing, text mining, & time series analysis for IT professionals and business decision-makers.

CONSULTANT | CAPGEMINI | 2012-2014

- Developed user classification models relying on decision trees, random forest and, logistic regression.
- Utilized analytical applications like R, Alteryx and Tableau to identify trends and relationships between different pieces of data and, draw appropriate conclusions.
- Spearheaded Customer Segmentation project using R, SQL and, Hive to analyze model results using cross validation, gradient boosting and ensembles.
- Gathered, analyzed, and executed projects using NoSQL, Python, and, Pandas.
- Served as a valuable knowledge resource to other data and analytics groups to make informed decision on how to tackle data mining problems.

ANALYST | EXCELLENCE SOLUTIONS | 2011-2012

- Performed regular analysis and reported on various forecasts and metrics such as business seasonality, market changes/conditions, inventory inflow and outflow, forecast accuracy, trending patterns, and correlations.
- Developed insightful briefs and recommendations on PowerPoint and Excel.
- Helped with the replacement of Excel reports with Tableau dashboards by using aggregated data stores created with SQL.
- Consulted with other teams to support data collection and integration using Excel, SPSS and SAS to support
- Used SPSS and SAS to develop ad hoc data analysis as requested by senior management.
- Created and maintained internal data cleaning tools using Java and SQL Server.

PROJECTS

Kaggle: Coupon Prediction (Multi-class Classification) – R, Tableau, Alteryx, SQLite

- Built decision trees, random forest and gradient boosted models to predict which type of coupon a customer is likely to buy
- Implemented ensemble – based classifier using decision trees, random forest and treebag to compare model performance

Amazon Web Services: Medical Billing Fraud Detection – Hadoop, SQLite, Java, Hive, Pig

- Find medical doctors who are using procedures not belonging to their specialty and bill Medicare
- Compared performance of standalone SQL RDBMS and Hadoop, Hive/Pig in pseudo, multi distributed mode with custom MapReduce implementation in Java

Advanced Data Analytics: Natural Segmentation occurring in Chicago's neighborhood – Weka, Python, R

- Developed an unbiased model of housing and home value in Chicago using Principal Components Analysis and Regression
- Generated heat maps to show the segmentation by combining the results of PCA, Regression and kmeans clustering