

Professional Summary:

- ✓ Having 8 years of total IT experience in Enterprise Project Development and implementations for providing solutions for Banking & Financial domains using **Java/J2EE, NoSql, Hadoop, Machine Learning** and **NLP**.
- ✓ Building predictive models using various Machine Learning techniques.
- ✓ Capable of processing and analyzing large amount of structured and unstructured data.
- ✓ Having experience in development of Hadoop Components using **Map reduce, Hive, Pig and Sqoop**.
- ✓ Worked and implemented in machine learning algorithms using **R** namely **Regression, Classification** and **Clustering Techniques**.
- ✓ Experience in Dimensionality reduction techniques.
- ✓ Experience in building ensemble models (Random Forest, Bagging and Boosting).
- ✓ Experience in handling class imbalance data.
- ✓ Experience in finding the entities and its relationship using NLP techniques.
- ✓ Good working experience in Data cleaning and Data Quality assessment.
- ✓ Worked and implemented CUSTOM UDFs in Pig and Hive.
- ✓ Expertise in Performance Tuning and Query Optimization using HIVE.
- ✓ Exposure to client interaction, user requirement analysis and Support.
- ✓ Experience in building Mappings for Data processing, Cleansing and Transformation.

Core Competencies:

Requirement Analysis, Data Wrangling, Anti Money Laundering, Data Visualization & Communication, Big data Analysis, Financial Crimes, Problem-solving skills, Exposure and Experience in various Big data tools

Educational Qualification:

- ✓ Post Graduation in **M.C.A** from **JNTU**

Organizational Experience:

- ✓ Bank of America. March 2014 to till today
- ✓ Neeha IT Services India Private Ltd. October 2012 to February 2014
- ✓ Nanomindz Technologies Private Ltd. June 2009 to September 2012

Areas of Expertise:

- **Big Data Technologies:** HDFS, Map Reduce, Hive, Sqoop, Pig
- **Machine Learning:** Regression, Classification, Clustering, Ensemble Modeling
- **Text Mining:** Sentiment analysis, Natural Language Processing
- **NoSql:** MongoDB
- **Version Control:** SVN
- **Programming:** Java/J2EE, R, SparkR

Activities:

- Actively participate the hackathons conducted by Kaggle and Analytics vidhya.
- Helping hand for other teams in data science applications within the bank.
- Contributing ideas for process improvement to save time and money for bank.
- Actively participate in data science meet-ups within & outside bank, Hyderabad.

Work Experience

Financial Crimes Analytics

Data Scientist

Bank of America

AML Analytics is for detecting and monitoring suspicious transactions in an effort to prevent money laundering. Collecting of data from all AML applications and push into HDFS. Load the data using data lake and run map reduce programs as per the business rules in order to fine tune the data and do fraud analytics on the customer and on different ways of transactions data. Using Clustering techniques analyze fraud and reduce the risk. Work collaboratively with Risk & Compliance during investigations of suspected and actual internal and external fraud. Review suspicious transactions for potential AML reporting and risk purposes and prepare regular reports to AML Compliance Officers.

Responsibilities:

- Extracted the data from AML applications and loaded into HDFS by applying business algorithms for fraud analysis.
- Worked on Exploratory Data Analysis for featuring engineering and Missing values imputation, Outlier detection to reduce the noise in data using R.
- Build and Monitoring the predictive Risk Models using various machine learning techniques.
- Validate and improve the performance of the model on unseen data.
- Analyze spending patterns, credit history, demographic data points to determine whether transactions are suspicious.
- Work close with detection channels team and reported to compliance.
- Collaborate closely with and provide Product Owners with business requirements that would improve risk and detection of emerging fraud trends.
- Monitor various fraud prevention/detection and suspicious activity alerts to identify potential risk and/or losses.
- Analyzes and investigates customer activity for potential AML and Fraud-related suspicious activity, through systemic alerts.
- Developed techniques to eliminate false alarms, estimate risks, and predict future of current transactions or users.
- Define and implement new strategies in fraud analysis.

Avail to Bank:

- Saves \$1.8M to the bank in the form of human interaction hours, infrastructure space and monitoring by using machine learning techniques in AML and keep improving to reduce the cost.

Customer Risk Assessment

Data Scientist

Bank of America

The Customer Risk Assessment(CRA) process is a risk-based methodology for assessing a customer's money laundering and sanctions risk. The CRA uses customer information gathered in the due diligence process in accordance with the GFCC customer identification and due diligence for Financial Crimes Compliance standards and ongoing customer's financial activity within the company to assign a risk rating. As part of the CRA process, compliance will identify customers engaged in particular types of activity as being Elevated or High Risk customers due to heightened financial crimes risk presented by the activity.

Responsibilities:

- Apply resampling techniques to avoid over fitting.
- Build classification model to predict the risk rating to the customer.
- Validate the model with various metrics.
- Applied boosting techniques to improve the performance of the model.
- Improving the behavior of the model on uneven CRA data.

Avail to Bank:

- Isolated the human interaction towards the CRA that save hours in terms of money.

Entity Recognition

Data Scientist

Bank of America

Crawl the data from social media and on web, clean the unstructured data using NLP techniques. Find out the entities from the data and its relations between the entities.

Responsibilities:

- Text preprocessing, Noise removal on unstructured data.
- N-Grams on sentence Tokenization, TF-IDF, Pos tagging on Tokenization of words.
- Topic modeling LDA on corpus.
- Find the entities and its relations on the corpus using NER.
- Text matching using cosine similarity.
- Text classification using a machine learning algorithm by training the entities.

OLMS

Big Data Engineer

Bank of America

Online List Management Services(OLMS) is a centralized application used internally by Bank Of America. This application is the main resource in providing watch list data to all applications within then bank irrespective line of business.

Responsibilities:

- Developed sqoop scripts to extract data form RDBMS to Hadoop and schedule the scripts.
- Developed hive scripts to handle different file formats and compression techniques.
- Automated process of creating hive tables across different layers (raw/work/stage/target) with different file formats.
- Write Map Reduce jobs and Pig script for applying business algorithms on the data for fraud analysis.
- Developed Pig scripts for distribution of data to downstream servers.
- Developed custom UDF for hive and pig for data validation and masking.
- Moved all crawl data flat files generated from various users to HDFS for further processing.
- The analyzed data exported to HDFS for data mining and data analytics.
- Optimization of complex hqls in hive by applying join strategy changes, bucketing, partitioning.

Gnaneswar Rao Mantri