

# Alok Yarnene

## Data Scientist - First Tennessee Bank

Piscataway, NJ - Email me on Indeed: [indeed.com/r/Alok-Yarnene/704741ef5016fc5f](https://www.indeed.com/r/Alok-Yarnene/704741ef5016fc5f)

- Over 6 years of experience in various domains like Banking, Healthcare, Insurance.
- Highly efficient in Machine Learning, Data Mining with large data sets of Structured and Unstructured data, Data Acquisition, Data validation, Predictive Modeling, and Data Visualization.
- Extensive programming skills in analytical and statistical programming languages such as Python 2.7, R 3.3.3, SAS 9.1, SQL.
- Experienced in Statistical Modelling and Machine Learning techniques like Linear and Logistic Regression, Decision Tree, Random Forest, Naïve Bayes, Neural Networks, Support Vector Machines and Cluster analysis and other machine learning algorithms.
- Experience with structured and unstructured data analysis using relational databases like Oracle SQL.
- Experience in handling big data on NoSQL databases like Cassandra 2.2, MongoDB 3.2.
- Adept in using Python libraries such as Pandas, Numpy, Seaborn, Matplotlib, Scikit-learn.
- Experience in using Jupyter notebook in Anaconda 3.0 an open source distribution for Python, R programming.
- Expertise in transforming business requirements into analytical models, designing algorithms, building models, developing data mining and reporting solutions that scales across massive volume of structured and unstructured data.
- Working knowledge in Big Data Topics like Hadoop 2.x, HDFS architecture, MapReduce.
- Expert in designing visualizations using, Power BI, Tableau 10.x software and publishing and presenting Dashboards, Storylines on web and desktop platforms.
- Strong experience in Software Development Life Cycle (SDLC) including Requirements Analysis, Design Specification in both Waterfall and Agile methodologies.
- Experience in using A/B test, Hypothesis test and ANOVA testing to find the accuracy of model.
- Knowledge in MS Project 2013 with phrasing projects, assigning resources, and phase durations.
- Well versed in Git repository for version control.
- Ability to independently and effectively organize and manage multiple assignments with excellent analytical and problem solving skills and ability to work independently and as a part of the team.

Willing to relocate: Anywhere

Authorized to work in the US for any employer

## WORK EXPERIENCE

### Data Scientist

First Tennessee Bank - Memphis, TN - August 2016 to Present

Title: Credit Risk Analysis (CRA)

First Tennessee, part of the First Horizon National Corporation, is a financial services company based in Memphis, Tennessee. It provides services to individual consumers and business of all sizes. Services include investments, insurance, loans and wealth management through its affiliate FTB advisors.inc.

Credit Risk Analysis (CRA) is a system used by First Tennessee bank to identify the risks of lending credits to loan applicants. Applicant data such as account status, credit history and employment history were collected and data pre-processing was applied to create a model that predicted that loan applicants have a decent credit. This information was then used by the Credit Risk team to quantify the risks associated with lending to applicants.

Responsibilities:

- Used Predictive Modeling, Statistics, Machine Learning, Data Mining, and other aspects of data analytics techniques to collect, explore, and extract insights from structured and unstructured data.
- Used Oracle SQL for data extraction from different sources.
- Performed Data Cleaning, Feature Scaling, and Feature engineering using Python 2.7 and Python libraries like Pandas 0.18 and Numpy 1.12.
- Implemented Support Vector Machine (SVM), Logistic regression model, K-means clustering for predictive analysis using python Scikit-Learn 0.18.
- Used Cross Validation to evaluate models for detecting Overfitting.
- Summarized the performance of models using Confusion matrix, Recall rate and Precision rate.
- Used A/B test and Hypothesis test to check the accuracy of the model.
- Visualized, interpreted, report findings and developed strategic uses of data using Power BI and Tableau 9.0, and created interactive Dashboards.
- Followed Agile methodology in entire project lifecycle and used Git for version control
- Documented all programs and procedures to ensure an accurate historical record of work completed on assigned project as well as to improve quality and efficiency.

Environment:

Python 2.7, Pandas 0.18, Numpy 1.12, Matplotlib 2.0, Sci-kit learn 0.18, ML lib, Anaconda 3.0, Oracle SQL, Power BI, Tableau 9.3, Agile, Git

## **Data Scientist**

HCA Healthcare - Nashville, TN - November 2015 to July 2016

Title: Clinical Data Management

Hospital Corporation of America is an American for-profit operator of health care facilities based in Nashville, Tennessee. They have multiple hospitals and surgical centers all around U.S

The Clinical Data Management system (CDM) deals with the development and implementation of HCA's Clinical Informatics Program as well as the expansion of business processes and performance related to HCA's Clinical Data Management (CDM) strategy. Data resources were used to improve the clinical and operating performance of HCA.

Responsibilities:

- Worked in all phases of research like Data Cleaning, Data Mining, Feature Engineering, Developing tools, Validation, Visualizations and performance monitoring.
- Handling huge data and performing creating, reading, updating and deleting(CRUD) operations on NOSQL databases like MongoDB 3.2.12.
- Used Python programming for managing and transforming and integrating with datasets in preparation for analytics.
- Performed exploratory data analysis like statistical calculation, data cleaning and data visualizations using Numpy, Pandas and Matplotlib.
- Developed predictive models using Logistic regression, Decision Tree, Random Forest and KNN algorithms.
- Used Cross Validation for checking overfitting in suggested model.
- To check the performance of model Confusion matrix, Recall Rate and Precision Rate.
- Performed Hypothesis testing to find the accuracy of the model created using machine learning algorithms.
- Created interactive Dashboards on desktop platform to visualize the data by using Power BI in MS Excel and Tableau 9.3.
- Project life cycle involved in Agile Methodology and used Git as version control.

- Interacted with the other departments to understand and identify data needs and requirements and work with other members of the IT organization to deliver and address those needs.

Environment:

Python, MongoDB, Oracle SQL, Numpy 1.12.1, Pandas 0.18.1, Scipy 0.19.0, Matplotlib 2.0.0, Sci-kit learn, Anaconda 3.0, Power BI, Tableau 9.0, Git

## **Data Analyst**

Logic Matter India Private Ltd - Hyderabad, ANDHRA PRADESH, IN - March 2014 to July 2015

Logic Matter, Inc is a software company that is based Hyderabad, India. They work on cloud business intelligence and big data services like Data Aggregation support, Analytics reporting support, and BI Solution engineering.

Project involved calculating bonuses for brokers depending on different types of compensation programs. This application loads large amounts of data from various sources. It calculates bonuses for all different states in the US, generates preliminary reports, statements, bonus summaries and Ancillary reports.

Responsibilities:

- Developed business process models using MS Visio 2013 to create case diagrams and flow diagrams to show flow of steps that are required.
- Designed, Implemented and automated modelling and analysis procedures on existing and experimentally created data
- Created PL/SQL packages and Database Triggers and developed user procedures and prepared user manuals for the new programs.
- Created dynamic linear models to perform trend analysis using Python 2.7.1
- Used MS Excel, MS Access and SQL to write and run various queries.
- Used traceability matrix to trace the requirements of the organization.
- Analyze the data and create dashboards using Tableau 9.0.
- Reviewed the logical model with application developer, ETL teams, DBA's and testing team to provide information about the data model and business requirements.
- Involved in the daily maintenance of the database that involved monitoring the daily run of the scripts as well as troubleshooting in the event of any errors in the entire process

Environment:

Python 2.7.1, SQL, Oracle 11g, MS Office, MS Visio, Tableau 9.0

## **PL/SQL developer**

Magna Quest Technologies - Hyderabad, ANDHRA PRADESH, IN - October 2012 to February 2014

Magna Quest Technologies limited is an enterprise solutions and products company based in Hyderabad, India. It provides solutions in areas of subscriber billing, and customer relationship management.

Worked on automation tool called SMART reports which generate reports showing the current trend of tickets, breach categories, count and related information of special tickets for the client IRCTC

Responsibilities:

- Requirement gathering from the Business Team.
- Analyze the Existing code and do the impact analysis.
- Performed DML and DDL operations in Oracle 9i as per business requirement
- Developed complex database objects like Stored Procedures, Functions, Packages and Triggers using Oracle Database, SQL and PL/SQL.

- Used SQL Loader to upload the information into the database and using UTL\_FILE packages write data to files.
- Involved in peer to peer code reviews.
- Code validation as per client requirement.

Environment:

Oracle 10g, SQL, PL/SQL

### **PL/SQL developer**

Birla Sun Life Insurance - Hyderabad, ANDHRA PRADESH, IN - April 2011 to September 2012

Birla Sun Life Insurance Company limited is life insurance based in Mumbai, India. It provides insurance services comprising protection solutions, children's future solution and others.

Bit soft is an insurance product for individual insurance where field agents attempt to go to individuals for getting their insurance plan registered. A database is created to store the data of the customers and their insurance plans.

Responsibilities:

- Worked on requirements gathering, analysis, design, change management and deployment.
- Automated the process of rebuilding indexes at regular interval for better performances.
- Generated reports that can handle both dynamic grouping and sorting.
- Created database objects like Tables, Views, Sequences, Synonyms, Stored Procedures, Indexes, Table Spaces and integrity constraints.
- Wrote SQL queries using joins, Sub queries and correlated sub queries to retrieve data from database.
- Debugged many PL/SQL packages procedures, function, cursors and types for application.
- Query optimization, execution plan and Performance tuning of queries for better performance.

Environment:

Oracle 10g, SQL, PL/SQL

### **SKILLS**

BI (3 years), business intelligence (3 years), Oracle (5 years), PL/SQL (4 years), SQL (5 years)

### **ADDITIONAL INFORMATION**

#### **TECHNICAL SKILLS:**

Languages\ Python Packages\

Python 2.7, R 3.3 programming, SAS 9.1, \ Numpy 1.12, Pandas 0.18, Matplotlib 2.0, \  
PL/SQL\ Seaborn 0.7, Scikit-learn 0.18\

Databases\ Big Data Tools\

Oracle 11g, MongoDB 3.2.12, Cassandra 2.2\ Hadoop 2.x, Hive 0.14, Spark 2.0\

Operating Systems: \ Microsoft Tools\

Windows 10/8/7/XP, Linux, Mac OS\ MS Office, MS Project 2013, MS Visio 2013\

Machine Learning\ Cloud Tools\

Linear and Logistic regression, Clustering, \ Anaconda 3.0 (Jupyter notebook), Horton \

SVM, Neural Networks, Decision Tree \ Works.\  
Random Forest, Naïve Bayes.\

Visualization Tools\  
Tableau 10.x, Power BI, Pivot Tables\