Tarun Balram
Data Scientist - Stage Stores Inc

Farmers Branch, TX - Email me on Indeed: indeed.com/r/Tarun-Balram/9d3c6242349dddf8

• Over 7+ years of experience in Machine Learning, Data mining with large data sets of Structured and Unstructured data, Data
Acquisition, Data Validation, Predictive modeling, Data Visualization.
• Proficient in statistical programming languages like R and Python including Big Data technologies like Hadoop, Hive with strong
domain expertise in Finance and Retail Sectors.
• Proficient in managing entire data science project life cycle and actively involved in all the phases of project life cycle including data
acquisition, data cleaning, data engineering, features scaling, features engineering, statistical modeling (decision trees, regression
models, neural networks, SVM, clustering), dimensionality reduction using Principal Component Analysis and Factor Analysis, testing
and validation using ROC plot, K- fold cross validation and data visualization.
• Extensive experience in Text Analytics, developing different Statistical Machine Learning, Data Mining solutions to various business
problems and generating data visualizations using R, Python and Tableau.
• 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 Hadoop, Spark, Hive and NOSQL databases like MongoDB and HBase.
• Hands on experience in implementing LDA, Naive Bayes and skilled in Random Forests, Decision Trees, Linear and Logistic
Regression, SVM, Clustering, neural networks, Principle Component Analysis and good knowledge on Recommender Systems.
• Good industry knowledge, analytical and problem-solving skills and ability to work well within a team as well as an individual.
• Highly creative, innovative, committed, intellectually curious, business savvy with good communication and interpersonal skills.
Willing to relocate: Anywhere
Authorized to work in the US for any employer

WORK EXPERIENCE

Data Scientist
Stage Stores Inc   -   Houston, TX - January 2017 to Present

Description:
Stage Stores, Inc is a retail company with 800 stores that provides apparel, accessories, cosmetics, footwear and housewares. People can buy it from stores or online. The project aimed to improve the stage.com recommender system by using market basket analysis to increase the sales.

Responsibilities:
• Implemented end-to-end systems for Data Analytics, Data Automation and Integration with custom visualization tools using R,

Mahout, Hadoop and MongoDB
• Performed Market-Basket Analysis and implemented decision trees, random forests, and K-fold cross validation
• Responsible for data identification, collection, exploration & cleaning for modeling, participated in model development
• Visualized, interpreted, report findings and developed strategic uses of data
• Responsible for using various machine learning algorithms - (Random Forest, SVM, Bagged CART)
• Interacted with the other departments to understand and identify data needs and requirements and work with other members of the IT
organization to deliver data visualization and reporting solutions to address those needs
• Analyzed data and provided reporting and insightful recommendations to business leaders on key performance metrics pertaining to sales & marketing
• Knowledge of other relational database platforms such as Oracle, MySQL, NoSQL, etc.
• Found Employee attrition rate using Decision Trees and Random forest
• Worked on Time Series forecasting using R
• Converted raw data to processed data by merging, finding outliers, errors, trends, missing values and distributions in the data
• Accomplished Data analysis, Statistical analysis, generated reports, listings and graphs

Environment: R, Python, Mahout, MySQL, Hadoop, MongoDB

Sr. Data Scientist
Roomer - New York, NY - February 2015 to December 2016

Project: Hotel Recommendation System

Description:
Roomertravel.com is a commercial website that provides Hotel rooms where people can sell their own hotel reservations.

The project aimed to improve Roomer Travel's Hotel Recommendation System by predicting a more personalized hotel recommendation for each user. The team was responsible for exploring streaming customer data, building and testing advanced Machine Learning algorithms and
Predictive Models, designing marketing experiments, and creating data visualization and web APIs.

Responsibilities:
• Participated in Data Acquisition with Data Engineer team to extract historical and real-time Data by using Spark
• Performed Data Mining and Data Cleaning using SparkSQL.
• Implemented Data Exploration to analyze patterns and to select features using Python.
• Built Factor Analysis and Cluster Analysis models using Python SciPy to classify hotels into different target groups and to identify
newly effective features.
• Built predictive models including Support Vector Machine, Decision Tree and Naïve Bayes Classifier using Python Scikit-Learn to predict the personalized hotel group for each user.
• Designed and implemented cross-validation and statistical tests including Hypothetical Testing, ANOVA, Auto-correlation to verify
the models' significance.
• Designed an A/B experiment for testing the business performance of the new recommendation system.

• Created reports and dashboards, by using Python Matplotlib and Tableau 9.x, to explain and communicate data insights, significant
features, models scores and performance of new recommendation system to both technical and business teams.
• Used Git 2.x for version control with Data Engineer team and Data Scientists colleagues.
• Used Agile methodology and SCRUM process for project developing.

Environment: Spark, MapReduce, HDFS, SparkSQL, Python 3.x, Matplotlib, Scikit-Learn, Tableau 9.x, SVM, Decision Tree, Naïve Bayes
Classifier, A/B experiment, Git 2.x, Agile/SCRUM

Finance Data Analyst
Penn National Insurance - Harrisburg, PA - July 2014 to January 2015

Description:
The aim of project was to analyze the revenue of financial data for the prediction of future trends and identify the risk associated with the capital
expenditures and worked with business users to understand their business functions and processes to help them define analytic and reporting
requirements, including data and metadata requirements for their systems that are complete and unambiguous.

Responsibilities:
• Review financial statements and analyze sales revenues, costs, expenses, and tax rates for all business units of leading process
automation company
• Developed detailed spreadsheets to identify trends and develop forecasts. Consolidate and analyze monthly results and budgets
• Performed ad-hoc analysis and reporting
• Performed extensive data quality analyses to evaluate existing data processes and recommended process fixes for accurately and effectively generate and measure data for analytics and reporting for leadership teams
• Analyzed and determined scope and volumetric parameters that dictate data profiling activity
• Constructed and executed complex SQL queries to verify business data rules and fit-for-use status of data under analysis
• Identified and incorporate required database performance objectives into established physical data models as part of data analysis and related specification process across multiple business areas of responsibility
• Performed extensive requirement analysis including SWOT, data analysis and GAP analysis
• Extracted data from existing data stores, Developed and executed departmental reports for performance and response purposes by using oracle SQL, R, procedures, packages, functions, database triggers

Environment: R, R Studio, MS Excel, Tableau

Business Analyst/Data Analyst
Marquette Savings Bank - Erie, PA - November 2013 to June 2014

Description:
The project was based on analysis of the business data and later providing the required strategies specified in unique business plans.
Responsibilities:
• Utilized skills in software applications such as R/Excel and analyzed data by decision tree analysis and regression analysis.

• Worked with business and technical staff to develop requirements document and specifications
• Responsible for extracting data from MS SQL Server, MS Access and Flat files, Data Warehousing & Database Design, ETL, Data
reporting and query, preparing data for analytics and project management
• Created database from representative project data and modified procedures, tables, views and constraints
• Generated reports for managers using dimensional modeling and reporting tools
• Edited raw data and created R data sets for statistical analysis for project/business decisions. Used R package for Statistical Analysis
• Created and maintained bulk data load & extract processes

Environment: R, Excel, MS SQL Server 2008

Software Business Analyst/Data Analyst
Geeta Infotech   -   Mumbai, Maharashtra   -   June 2010 to October 2013

Description:
The project was web-based College Management System where users can share information about grades, attendance records and homework
assignments etc. Four different types of data were processed and analyzed such as Administrators, Students, Teachers and Parents.

Responsibilities:
• Worked on different tasks of data analytics pipeline like data cleaning/transformation, and exploratory analysis of large data sets
using SAS and SQL
• Responsible for the design and development of Business Intelligence (BI) reports
• Responsible for communicating functional requirements to the entire team and training new colleagues of the project
• Analyzed accuracy of the validity and accuracy of Data using CSV files
• Actively participated in Waterfall model development
• Created diagrams and documentation for team resources, company reference, and client usage
• Collaborated with management and product owners to maintain coordination between products
• Facilitated meetings for consistent and timely communication and issue resolution

Environment: SAS, SQL, MS Excel

EDUCATION

Bachelor of Engineer in Computer Engineering in Engineer
Mumbai University   -   Mumbai, Maharashtra

SKILLS

algorithms (2 years), Data Cleaning (5 years), Excel (4 years), machine learning (2 years), SQL (4 years)

ADDITIONAL INFORMATION

COMPUTER SKILLS
Software: Visual Studio 10, MS SQL Server 2008, Erwin, Microsoft Excel & Access, Tableau
Languages: R, C, C++, Java, Python, Java Script, PL/SQL, SQL, VB.Net

Concepts: Data Structures, Data Cleaning, Data Mining, Machine Learning Algorithms, Statistical Modelling, ETL Process, SSIS, Big
Data, Hadoop, NoSQL, Mahout, Amazon EC2, Hive, Spark, MongoDB, HBase, Relational Database, Oracle
Packages: Python: Pandas, NumPy, SciPy, SymPy, Scikit-Learn, Matplotlib; R: ggplot2, gmodels, R Notebook; HiveSQL, SparkMlib,
SparkSQL