Karthik C

Lead Data Scientist - Best Buy

Minneapolis, MN - Email me on Indeed: indeed.com/r/Karthik-C/8235108e51fa18a0

• Data Science professional with 6+ years of experience in delivering end to end data science projects using R, SAS, Python, Hadoop, Spark, Tableau and SQL,pySPARK.
• Proficiency in statistics, machine learning and creating compelling visualizations.
• Worked in versatile roles for Retail, Banking, Insurance and analytics product development.
• Extensive exposure on analytics project life cycle CRISP-DM (Business understanding, Data understanding, Data preparation, Modelling, Evaluation and Deployment).
• Experience on data wrangling, data visualization and reporting using R, SAS, Tableau and Alteryx.
• Well versed with data preparation, data cleansing, descriptive statistics reports using R packages and SAS.
• Profound analytical and problem solving skills along with ability to understand current business processes and implement effective solutions.
• Experienced in loading and manipulating the data from HDFS/data lake to Apache Spark.
• Proven to be curious to mine hidden insights and present complex data analytics to non-analytic audience.
Willing to relocate: Anywhere
Authorized to work in the US for any employer

WORK EXPERIENCE

Lead Data Scientist

Best Buy - April 2016 to Present

• Developed a supply chain model based on a many-to-many directed graph, expressed as a mixed integer linear program, to determine the optimal inventory flows which minimize transportation and processing costs given initial conditions for demand and inventory positioning.
• Used data from an open map and routing service to investigate strategic service areas in key markets based on drive-time and -distance.
• Lead the technical development of a Python+Teradata application to identify the root causes across the supply chain at item-location-date resolution and scaled to production-class workloads.
• Collaborated in the improvement and calibration of an inventory flow/replenishment model, as well as statistical analysis of the model results to provide recommendations for optimized distribution strategies for small-format stores (formerly Target Express).
• Developed a volume-based pallet count estimation method based on distribution invoice data and known carton dimensions to help support the growth and operations of small-format stores.

Data Analytics Lead

United Health Group - August 2015 to April 2016

• Generated a classification models based on claims data to determine the probability of any individual being claims rejected in future and final model was 89% accuracy in multi-fold validation checks.
• Analysed accuracy of data acquisition and standardization via statistical methods
• Visualized client claims data using Excel, Tableau and R.
• Executed on research/analyses to provide insight into client needs
• Performed appropriate research/analysis to drive product strategies (e.g. product characteristics, functionality, root cause analysis, release strategy)
• Engaged with project sponsors and stakeholders to understand business problem.
• Engaged in bringing the structure to each request and translate requirements into an analytic approach.

• Spark SQL uses the Spark engine to execute SQL queries on data sets persisted in HDFS or data lake.
• Involved in converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala
• Load the data into Spark data frame and do in memory data Computation to generate the Output response.
• Import the data from different sources like HDFS/data lake into Spark.
• Handled importing other enterprise data from different data sources into HDFS using Sqoop and performing transformations using Hive, Spark.
Email Campaign Marketing:
• Business Objective: Task was to build a model to optimize the marketing efforts of a company, by predicting whether customer will respond to the marketing campaign and purchase a policy or not. The main objective here is to predict potential customers, rather than classifying customers and non-customers accurately.
• Implemented Logistic Regression, c5.0 boosting and random forest algorithms. Random Forest gave best recall value and could predict potential customers more than any other models. It predicted least False Negatives and more True positives, supporting our assumption.
• TOMEK Links and SMOTE algorithms are applied to overcome this unbalanced data issue.
• Assumed that the cost associated with email marketing is very less and profit by predicting one extra potential customer is more.
• Model had low accuracy but had high predictive power. Trade-off between accuracy and recall is implemented for optimizing.

Data Scientist

Kohl's - August 2013 to August 2014

• Predicted sales of a store given the store attributes to plan budgeting, staffing and marketing.
• Built an ensemble of machine learning models to forecast sales and perform model comparison using MAPE.
• Segmented the customers based on demographics using K-means Clustering.
• Generated insights from the clusters to perform target marketing based on the customer profile such as age, education, income levels and median rooms.
• Developed data pre-processing modules and rule extraction engines in R using Random Forest and Decision Trees for an analytics product. Reduced the execution time by 30%.
• Performed data discovery by generating compliance trend reports to show the calculated value for certain compliance rules over time for significant KPI's.
• Developed adhoc reports on top performing products on a weekly basis using tableau.
• Develop campaign insights from a combination of proprietary client data, syndicated industry information, with geographically relevant consumer and vendor demographics.
• Audited and validated the customer selection results - segmented based on behaviour, predictive models, and other criteria.
• Supported new products, communication methods, marketing campaigns, and brands.
• Created visualizations, reports and dashboards to present clients and stakeholders.

Senior Data Analyst

Wells Fargo - June 2011 to August 2013

• Developed visualizations and dashboards of insurance claims using Tableau to extract significant patterns.
• Analyzed large volumes of data to draw actionable insights that supports business decision making process.
• Generated periodical reports using SQL and transformed large amount of data as per the business requirement.
• Developed and tested hypotheses to check whether the results are statistically significant or not.
• Developed predictive models such as logistic regression, Decision trees, Random Forest to classify whether to approve a loan to customer based on credit history and demographics.
• Provided Root Cause Analysis by understanding the process flow, and implemented code fix for the issues.

EDUCATION

Masters of Science in Data Science in Data Science

Ohio State University
August 2015

Certification

Indian Institute of Management
August 2014

Bachelor of Engineering in Electronics and Communication in Electronics and Communication

Anna University
April 2011

SKILLS

Python (3 years), random forest (3 years), SAS (3 years), Serial Attached SCSI (3 years), SQL (3 years)

ADDITIONAL INFORMATION

TECHNICAL SKILLS

Languages R, SAS, Spark R, Java, Android, Python, SQL, Spark
Machine Learning
K-NN, Decision Tree, Naïve Bayes, Random Forest, K-Means clustering, Gradient Boosting, Market Basket Analysis, RFM Analysis, Neural Networks, Principal Component Analysis

Statistics Linear Regression, Logistic Regression, Time Series
Data Visualization Tableau, QlikView, R-shiny
Big Data Hive, Pig, Sqoop, Map Reduce, Hadoop
Tools Excel, Pivot Table, SSRS, SSIS, Cognos, Alteryx, MS Access, Google Analytics