Vinay Karagod

Data Scientist

Wichita, KS - Email me on Indeed: indeed.com/r/Vinay-Karagod/22d7fc634b127c74

Willing to relocate: Anywhere
Authorized to work in the US for any employer

WORK EXPERIENCE

Data Scientist

H3 Technologies - February 2017 to Present

Description: Involved in various projects such as Sales & Marketing analytics, inventory optimization for Global Industrial companies with a wide presence in North America, Client clustering and Resource optimization projects using both R, Python in Anaconda, Spark (PySpark) and Revolution R for big data projects.

• Designing and developing various machine learning frameworks using python, R, and Matlab.
• Collaborated with data engineers to implement ETL process, wrote and optimized SQL queries to perform data extraction from Cloud and merging from Oracle 12c.
• Collected unstructured data from MongoDB 3.3 and completed data aggregation.
• Performed data integrity checks, data cleaning, exploratory analysis and feature engineer using R 3.4.0.
• Conducted analysis on assessing customer consuming behaviors and discover value of customers with RMF analysis; applied customer segmentation with clustering algorithms such as K-Means Clustering and Hierarchical Clustering.
• Developed personalized products recommendation with Machine Learning algorithms, including Collaborative filtering and Gradient Boosting Tree, to better meet the needs of existing customers and acquire new customers.
• Used Python 3.X (numpy, scipy, pandas, scikit-learn, seaborn, NLTK) and Spark 1.6 / 2.0 (PySpark, MLlib) to develop variety of models and algorithms for analytic purposes.
• Coordinated the execution of A/B tests to measure the effectiveness of personalized recommendation system.
• Performed data visualization with Tableau10.0, MeteorJS and generated dashboards to present the findings.
• Recommended and evaluated marketing approaches based on quality analytics of customer consuming behavior.
• Determined customer satisfaction and helped enhance customer experience using NLP.
• Used Git 2.X to apply version control. Tracked changes in files and coordinated work on the files among multiple team members.

Data Scientist

Wichita State University - August 2015 to December 2016

1.My role was to perform statistical analysis for industrial market research data collected from surveys, give insights into specific business problems, preparing technical documentation and present results in a simple and clear fashion to a non-statistical audience.
• Identified risk level and eligibility of new insurance applicants with Machine Learning algorithms.
• Predicted the claim severity to understand future loss and ranked importance of features.
• Used R 3.X, R2.X and Spark 1.4 (PySpark, MLlib) to implement different machine learning algorithms including Generalized Linear Model, SVM, Random Forest, Boosting and Neural Network.

• Evaluated and optimized performance of models, tuned parameters with K-Fold Cross Validation.
• Provided analytical support to underwriting and pricing by preparing and analyzing data to be used in auctorial calculations
• Designed dashboards with Tableau 9.2 and MeteorJS provided complex reports, including summaries, charts, and graphs to interpret findings to team and stakeholders.
• Identified process improvements that significantly reduce workloads or improve quality.
• Utilized SQL and HiveQL to query, manipulate data from variety data sources including Oracle 10g and HDFS, while maintaining data integrity.
• Worked on data cleaning, data preparation and feature engineering with Python 3.X including Numpy, Scipy, Pandas, Matplotlib, Seaborn and Scikit-learn.
• Worked with the version control tools, such as Git 2.X, to keep versions attributed from different people and record project at different time points.
• A Novel Machine Learning Framework For Phenotype Prediction Based On Genome-Wide DNA Methylation Data. [Published IJCNN 2017]
• Recognizing handwritten digits using artificial neural network
Implemented a project using python, skit-learn, Numpy, Scipy and UNIX for detecting handwritten digits using artificial neural networks (Natural Language Processing) and convolutional neural networks.
• Library Management system
Developed a database using JAVA, SQL, and UNIX for managing a library, with one user as a librarian with limited access and the other as a database administrator.

Data Scientist

PassFail.com - Wichita, KS - May 2015 to July 2015

Description: Developed python scripts in Anaconda for variation and Time Series analysis, of Stock Market, and for Rent updates analysis from passfail.com

• Prediction of Multiple Regression analysis.
• Designed and developed an advanced recommendation system for real estate and finance customers utilizing R and Python.
• Utilized SQL to extract data from SQL Server 11.0, and MongoDB, to prepare data for analysis.
• classified customers by RFM analysis, clustering and regression model with R 3.0, selected customers with high value and improved their retention rate by sending ads and coupons

Environment:
Excel 2013, R 3.0, Hadoop 1.X, MS SQL Server 2012, Tableau 8.1

Data Scientist

Tech Mahindra Inc - Bangalore, Karnataka - April 2012 to July 2014

Description: My role was to design and implement A/B Testing for British Telecom, capture data using web analytics using web crawling platform and Optimization, create dashboards using ELK (Elastic search, Logstash and Kibana) framework.

• Used Python 2.7 to apply time series models, clustering algorithm and other data mining methods to explore the fast growth opportunities of our clients
• Analyzed the traffic queries of Baidu search engine using classification algorithm.
• Assisted to improve the liquidity of our ads model.
• Based on the data of clients and traffic, designed comprehensive analysis to optimize products and explored the strengths and weaknesses of products.
• Team Member, Rating Engine Application

Boosted marketing for client by designing a cell phone data rating engine in Python, PL/SQL, and UNIX/Linux to suggest suitable cell phone plans based on data usage.
• Team Lead, Call Center Application
Increased efficiency of client customer support by implementing a call center application using R, Python and SQL that tracked customer complaint history and assigned defects accordingly.
• Team Member, Open Reach
Enhanced customer buying experience using Machine learning, R, python and SQL
• Team Member/SPOC [R50 Release], One Siebel
Ensured smooth access for customers by developing scripts in Python and Unix for managing the crash reports and log levels of Unix and Windows servers, maintaining server network configuration components using Siebel 8.1 and Oracle, analyzing and providing RCA for server defects, and developing and publishing documents for avoiding defects in future releases; set up Scrum daily between Delivery Managers, Dev Team, Test Team, and Admin Team.

Environment:
Excel 2010, Python 2.7, Hadoop 1.X, Mapreduce, MS SQL Server 2008, Tableau 8.0

EDUCATION

M.S. in Computer Science
Wichita State University - Wichita, KS
December 2016

B.S. in Electronics and Communication Engineering
Rao Bahadur Y.Mahabaleshwarappa Engineering College
May 2011

PUBLICATIONS

A Novel Machine Learning Framework For Phenotype Prediction Based On Genome-Wide DNA Methylation Data
February 2017
DNA methylation (DNAm) is an epigenetic mechanism used by cells to control gene expression, and identification of DNAm biomarkers can assist in early diagnosis of cancer. Identification of these biomarkers can be done using CpG (Cytosine-phosphate guanine) sites, or particular regions in DNA. Previous machine learning methods known as MS-SPCA and EVORA have been used to link DNAm biomarkers to specific stages of cervical cancer using CpG data. In this paper, we propose a novel machine learning framework that yields greater AUC accuracy than the MS-SPCA and EVORA for predicting stages of cervical cancer using CpG data. This framework appears to be promising in regards to the data examined herein as well as for future biological studies.

ADDITIONAL INFORMATION

Technical Proficiency:

• Machine Learning and Deep learning Techniques: Trees, Bayes Model, SVM, Ensemble Methods, Neural Networks, RNN, KNN, CNN, MLP, Ensemble SVM, Majority voting, Linear models, Classification, Regression, Clustering, Kernel methods, Memory Networks, LSTMs, Dimension reduction, Deep belief networks, Time series Models, Statistical tests Computer vision, Reinforcement learning.
• Programming Languages: C, C++, SQL, PL/SQL, JAVA, UNIX, XML, SIEBEL 8.1 CCC, HTML5, CSS3, python, Matlab, R.
• Tools: Toad, SQL Developer, SQL Server, SQL Plus, putty,TensorFlow,Torch,Caffe,Lua, HP Quality Center, Eclipse,