DEEPESH KUMAR

14415 D102 Hellenic Drive, Tampa FL   (813)-606-9570   deepesh@mail.usf.edu   Linkedin

EDUCATION:

University of South Florida Muma College of Business, Tampa, USA   **Expected Graduation: DEC 2017**

Master of Science, Business Analytics and Information System   GPA: 3.9

Coursework:

Data Mining	Data Warehouse	Operation and Supply Chain-Management
Statistical Data Mining	Big Data	
Advanced Database Management	Statistical Programming(SAS)	

University of Mumbai, India   **JUNE 2013**

Bachelor of Engineering, Computer Science Engineering (First Class with Distinction)   GPA: 3.6

Technical Skills:

Database Skills: Oracle (9i, 10g, 11g), MySQL, Replication, Migration, Dimension Modelling, NoSQL- Cassandra, MongoDB

Big Data: Hadoop, HDFS, Hive, Pig, HBase, MapReduce, Spark, sqoop, flume, kafka, DynamoDB

Data Mining Tools: SAS Enterprise Miner, RapidMiner, Qlikview, Google Analytics & Ad Words, Marketing Analytics, NLP

ETL & Reporting: Tableau, Qlikview, Pentaho, Spotfire, RStudio, Anaconda-Python, IBM SPSS, Talend, Informatica

Programming Languages:

Proficient: Java, J2EE, Python, R, SAS, Oracle PL/SQL, Scala   **Familiar with**: C#, C++, Asp .Net MVC 5, VB .Net

Work Experience:

Software Engineer, Goldensource India Pvt. Ltd., India   **OCT 2013 – JUL 2016**

Implemented Data warehouse and Goldensource Orchestrator(ETL) for our client World Bank.

- Mapping of Goldensource Data Model with Business Feeds from Bloomberg, Reuters and IDC
- Used Java, J2EE, Hibernate, Spring for Workflow Activity Development and manage daily functional checklist with Agile software development.
- Experienced in data migration of tables, views, functions, scheduled jobs to virtualized platform
- Wrote Stored procedures and triggers for transformation(ETL) logic and materialized view for report generation.
- Automated the data load process from various feed in staging area to the data warehouse, reducing man hours by 70%.
- Expertise in Trouble shooting and Analysis of millions of Trades, Cashflows and Risk transactions on a daily basis.
- Analyzed database server logs to recommend solutions for optimal storage configuration and associated risks.
- Developed and tested disaster recovery plans such as clustering, replication, for client's huge production databases

Academic Project:

Human Activity Recognition using Machine Learning in R (Data Mining)   **OCT 2016 – NOV 2016**

The data was collected from accelerometers on the belt, forearm, arm, and dumbbell of 6 participants. They were asked to perform barbell lifts correctly and incorrectly in 5 different ways.

- Predicted the manner in which they did the exercise. Best predictive model generated using random-forest technique and tested its performance on cross validation dataset.
- Project URL: https://deepeshkashyup.github.io/Practical-Machine-Learning/Practical_Machine_Learning_Project.html

Regression on Used Car Data (CarFax) (Python, SFrames, Scikit-learn, Matplotlib)   **OCT 2016 – NOV 2016**

Determined various trends in used Car purchases using Exploratory Data Analysis, Hypothesis Testing and Logistic Regression.

- Scraped Real time data for cars from CarFax using Web Crawler written in python using Beautiful Soup and Requests APIs.
- Performed data cleaning, data transformation, data evaluation and analysis using Python packages.
- Used regression models to predict car prices with a root mean square error of 27$ and accuracy of 85% for a cut off $1000.
- Project URL: https://github.com/DeepeshKashyup/WebScrapperKbb

Exploratory Data Analysis of Indian Cities using Shiny R (Statistical Data Mining)   **JAN 2017**

Built a reactive web application in Shiny R which explored the story of Indian cities with the help of Maps, Plots and Tree Maps.

- Project URL: https://deepesh.shinyapps.io/sdmapp/

Machine Learning Algorithm: Classification-Analyzing Sentiment (Python, SFrames, Scikit-learn, Matplotlib)   **DEC 2016 – FEB 2017**

- Loaded and explored product review data, explored most popular product using visualization in Python, defined which review has positive or negative sentiment, evaluated a classifier and ROC curve.
- Analyzed data sets by regression, classification, and clustering, assessed model quality by error metrics.

Certifications

- Coursera Certified: R Programming, Statistical Inference, Practical Machine Learning, Python, Regression Models
- Big Data University Certificates : Hadoop, Spark
- Citigroup/USF Third Party Risk Management Certificate
- Microsoft Certified Professional in C#.Net Web Development

Achievement/Awards:

- Coursera Mentor for Regression Models.
- Awarded the Associate of the year 2014 in GoldenSource Corporation, Mumbai.