Ravi Shankar

Contact Info:

Phone: 3377391082
e-mail: ravi_ladhar@live.com
Address: 800 Pleasant Street, Apt # 1404, New Bedford MA
LinkedIn: https://www.linkedin.com/in/raviladhar/
GitHub: https://github.com/raviladhar

A dedicated person with a master's degree in computer and information science from the University of Massachusetts Dartmouth and 2 years of experience in data analysis, statistics, data visualization, SQL, SSRS, and machine learning. Recently I completed my master thesis in geospatial data analysis, consists of building different statistical techniques to focus more on insights through different trends and visualizations

Technical Skills:

Programming Languages: Python (Pandas, NumPy, Scikit-learn, Matplotlib, Plotly, Folium, MapboxGl, PySpark), R, C#, SQL, HTML5, CSS, JavaScript (D3.js, JQuery, MapboxGL)
Tools/Technologies: Tableau, Spotfire, AWS (S3, EMR), Redshift Amazon, MySQL, SQL Server 2015, AWS, Visual studio 2015, Java Net Beans IDE, Weka data mining tool, Anaconda 3.0 (Jupyter Notebook, Spyder), RStudio, Latex, Linux OS (Ubuntu), Windows, VMware virtual player.

Research Experience:

Topic: Analyzing Noise in Trajectory Data — Sept 2017 – May 2019
University of Massachusetts Dartmouth, MA

Researched and developed new techniques for trajectory data cleaning process, where data contained multiple outliers, bad segments, time and distance gaps. Used python for data wrangling, pre-processing, visualization and outlier detection techniques.

- Used statistical and mathematical concepts to build centered Z-Score technique on speed and perpendicular distance to detect and remove outliers. Contributed the new segmentation technique to break the trajectories and remove large time intervals gaps.
- The results effectively represent the cleaned trajectory data, where these results were visualized through multiple line charts, graphs and mapbox-gl maps.
- Created maps using Geojson (geometry) files to visualize a large amount of data
- Used Python libraries and JavaScript (MapboxGL)
- Link to the code: https://bit.ly/2EPtRGH and Presentation: https://bit.ly/2WzR5Lb

Datasets: Applied these proposed techniques on two different datasets, these datasets are "Vessels Traffic Data" which contains multiple vessels information and "Geo Life Trajectory Dataset" which contains taxi data

Professional Experience:

Software Developer — Mar 2014 – Dec 2015
Bank Al Habib Limited Karachi

Worked on different banking modules such as account opening, banking requests and authentications for the higher-level approvals, and other multiple modules using C# MVC 4 framework with SQL server database management queries, procedures and multiple script writing.

- Build SQL queries, created tables, views, pivot tables and stored procedures.
- Integrated SQL Server Reporting Services (SSRS) tool, built SSRS reports through SQL Procedures, complex queries, and performed business calculations to generate monthly and yearly banking reports with large amount of data
- Worked on Microsoft Team Foundation Service tool embedded with Visual Studio for bugs reporting and resolving
- Practiced agile methodology for software performance and quality assessment
- Resolved multiple bugs in the system for the quality of the system

Ravi Shankar

Internship
Pakistan Telecommunication Limited (PTCL) June 2012 – Aug 2012

Built a module for the IP-Core services to record and access different IP and sub-netting information. Migrated multiple excel sheets to MySQL database, and used PHP to build and design their internal webpage for Create, update, Read and Delete CURD operations.

- Developed a web based internal system to put their data at one place, for querying easily
- Built a dashboard which supports insert, update, search, report generation facilities
- Completed the assignments of calculation of sub-netting

Projects:

Santander customer transaction prediction (Machine Learning project)

- Worked on machine learning algorithms to predict the customers transaction
- Used Logistic regression, Decision tree and Random Forest to train and test the model on the data. and used the confusion matrix to describe the performance of the model.
- Data was unbalanced, used SMOTE to balance the data equally and then to train the model
- Used Kaggle data for the analysis
- Link to the code and results https://bit.ly/2N11Fr5

New York City Bikes data analysis (Data Visualization project)

- Worked on some interesting questions to obtain insights from the data through effective visualizations
- Data obtained from NewYorkCityBikes webpage, which required the data preprocessing like data cleaning, filtering and merging, etc.
- Visualized and analyzed some important trends and fact about the bike rides based on their station locations traffic analysis based on different times of the day, bikes rides based on the ages, and trends dependent on number of rides and weather
- Used R to preprocess the data, JavaScript (D3.js) and Leaflet for the map library
- Link to the demo, code, preprocessed data and report: https://bit.ly/2LLTHkP

Trajectory data Analysis (Scalable Data Analysis)

- The goal of this project was to compare the processing power of two python libraries (Pandas VS PySpark), where large amount of data (9.00 GB) was used
- Based on the results, PySpark performed very well in-term of computation and results, whereas Pandas library took much more time to preprocess the same amount of data
- Used multiple analysis using Trajectories data

ERP System – Home Garden Shopping (PHP and MySQL)

- Designed and developed an ERP System where user can order online different gardening tools
- Used multiple CURD operations, online requests and order systems, viewing items and their descriptions
- Used PHP and MySQL with different JavaScript functions

Education:

- Master of Science in Computer and Science Sept 2016 – May 2019
  University of Massachusetts Dartmouth, MA, USA.

- Bachelor of Science, Computer Science Aug 2009 – May 2013
  Shaheed Zulifkar Ali Bhutto Institute of Science and Technology (SZABIST) Karachi.

`