Aadish Chopra

aadishchopra16@gmail.com | 3128689603 | Chicago, IL

SUMMARY

Experience of 3 years in data analytics, statistical analysis, predictive modelling, and proficiency in AWS and SQL and in developing visualization reports using Tableau, RShiny and PowerBI

EDUCATION

Master of Science in Management Information Systems Dec 2017
University of Illinois at Chicago – Liautaud Graduate School of Business
Bachelor of Technology in Electronics and Comm. - Rajasthan Technical University May 2015

TECHNICAL SKILLS

Languages: R, SQL, Python, Java, C++, JavaScript, Bash, Linux Shell Scripting, VBA, PowerShell, DAX

Tools: Excel, SQL Server, PostgreSQL, Tableau, Power BI, Kubernetes, Docker

Big Data: Redshift, AWS (S3, EC2, Lambda), Spark, Hadoop, Hive, Azure, SAP HANA

Machine Learning: Random Forest, XGBoost, SVM, KNN, Regression, ARIMA, HMM, Latent Dirichlet Allocation, Bayesian Networks, Meta-Analysis, TensorFlow, mxnet, keras, torch

EXPERIENCE

Software Developer 2: ORACLE Jan 2020-Present

Data Analytics Consultant: DISYS Nov 2018 – Dec 2019

- Development and automation of Claims Analysis Tool using PowerShell and SharePoint REST API. RECAT Tool allows clients to get a quick summary of their claims. Improved utilization of resources of upto 2 FTE and saved $24,000 on server purchase.
- Developed tree based predictive models for grading real estate properties and for low premium machinery and crop insurance
- Built dashboards like Utilization Management and Demand Modelling using SSRS, FlexDashboard (R) and PowerBI.
- Created and scheduled Spark ETL jobs using AWS GLUE for data ingestion to Redshift for creating business specific reports.
- Created a document management system using .NET framework and RCurl. Reduced report building time by 300%

Business Intelligence Engineer: CENTENE (for Insight Global) May 2018 -Nov 2018

- Predictive modelling on usage of state Medicaid funds using regression (lasso, ridge, polynomial and splines)
- Fetch nearby hospitals data through Google maps API using python and used pandas to structure data
- Created stored procedures and queried SQL server database for reports hosted on SSRS
- Text mining (word2vec, tf-idf) on IVR transcripts to identify member satisfaction, customer drop-out rate

Data Analyst (INTERN): NIELSEN CATALINA SOLUTIONS (for Adecco) Oct 2017 - Dec 2017

- Created data pipelines with Unix for data analysis and migrated those batches to Hadoop
- Worked on campaign studies to find sales lift. Used neural nets to find growth factors in advertisement sector
- Mapped Home scan data from Nielsen to frequent shopper data from Catalina

Data Scientist (INTERN)**:** LOYOLA MEDICAL CENTER June 2017- Aug 2017

- Worked on Ordinary Differential Equations, stepwise regression for Hepatitis B and Hepatitis D virus, explaining the pharmacokinetics of drugs and virus interactions
- Analyzed influential factors that affect treatment outcomes by fitting Logistic Regression Models in different age groups
- Applied Biphasic Models to fit viral kinetics to predict their treatment outcomes, time to cure and analyze accuracy

Data Engineer: TATA CONSULTANCY SERVICES Dec 2015- June2016

- Analyzed data using SAP HANA by building calculated view, attribute view constraining the query's response time and migrated data from ORACLE to SAP HANA. Executed SQL queries for reporting/dashboard
- Created a website using Java, Angular JS, Hibernate and POSTGRES for a Healthcare company

PROJECTS Github

NATURAL LANGUAGE PROCESSING (HEALTH TWEETS)

- Built a Web scraper in RSelenium to extract information like tweets, reposts, followers, following, likes etc.
- Cleaned tweets using tm package of R and using regex expressions
- Classified tweets using KNN and Latent Dirichlet Allocation in Alerts, General Information, Promotion etc.

HADOOP (MWRD, CHICAGO)

- Developed a 2-node cluster using Cloudera Hadoop for Metropolitan Water Reclamation District, Chicago, IL
- Created Spark ETL Jobs on Hadoop Cluster for data transfer using Talend Cloud

- Key Performance Indicators were calculated using exploratory data analysis, visualization and tf-idf analysis

AMERICAN HEALTH ASSOCIATION:

- Built statistical model to predict level of health information exchange among hospitals across United States
- Statistical techniques such as PCA, Discriminant analysis used to decompose the survey and reduce variables

LEANTAAS:

- Calculated the utilization of operating theatres using the transaction logs
- Analyzed requests, transfers and releases from the transactions log and calculating denial rate and response time
- Found usage patterns and suggesting improvement in the current scheduling system