VIPIN DAS

Vancouver, Canada +12368632533 vdas@sfu.ca vipin-das-09875037
https://placemedas.github.io/ https://github.com/placemedas

SUMMARY

Data Engineer with 9 years of industry experience in building data intensive applications, tackling challenging architectural and scalability problems in Finance and Automotive sectors. Currently a graduate student at Simon Fraser University(SFU) specializing in Big Data holding a CGPA of 4.00/4.33.

SKILLS

- Languages: Python,R,SQL,Cobol,Java
- Databases: Cassandra, IBM DB2,Dynamo DB
- Visualization: SAP Analytics Cloud, Tableau
- Web Framework: Plotly Dash ,HTML, Flask
- Machine Learning : Linear and Logistic Regression, Clustering, Decision Trees , Neural Networks
- Big Data Technologies: Apache Spark,Hadoop,Kafka
- Cloud Technologies : GCP
- Data Analysis: Pandas, Numpy, BeautifulSoup
- Agile Tools: Microsoft TFS,JIRA
- Version Control: Git
- Transferable : Teamwork , Leadership

PROFESSIONAL EXPERIENCE

Analytics Intern | SAP Inc. ,Canada **May 2020 – Present**
SAP Analytics Cloud

- Developed various KPI dashboards using SAP Analytics Cloud to analyze the operational data and provided recommendations to improve the quality of product.
- Created a data pipeline to collect, preprocess, transform and visualize skill development data using (MongoDB and SAC) resulting in improved and customized training plan for employees.
- Identified gaps in hiring process and improved gender diversity ratio to a level of 60:40 by performing exploratory data analysis on HR data.
- Introduced a chatbot using SAP Conversational AI, Python web scrapper and Flask web engine to interact with support teams to quickly discover 150+ on-call contacts across different time-zones.

Teaching Assistant | Simon Fraser University ,Canada **Jan 2020 – Apr 2020**
CMPT354 – Database Systems

- Assisted 114 students by conducting weekly tutorials on database concepts as well as aided professor in proof reading and correction of 5 assignments.

Application Development Specialist | Volvo Group Ltd., India **Mar 2018 – Aug 2019**
Data Engineering and Administration

- Performed data analysis and demand forecasting for Volvo parts using time series forecasting methods such as ARIMA and Exponential Smoothing to reduce warehouse wastage by 30%.
- Enhanced and maintained master data ETL frameworks and interfaces for entire Volvo Group using SQL,DB2, COBOL and REXX ensuring continuous operation across all Volvo brands with zero defects.
- Implemented a streaming environment using IBM MQ and IMS DC to support REST API services such as request-reply and publish-subscribe for 10000+ users.

Module Lead/Business Systems Analyst | Mindtree Ltd., India **Nov 2013 – Mar 2018**
American Express Global Business Services

- Elicited business requirements and engaged with multiple stakeholders to integrate IRIS Optical character recognition into procure to pay(P2P) operations resulting in 100% digitization of invoice process.
- Designed COBOL-XML programs and wrote SQL queries to fetch, join and aggregate data from multiple IBM DB2 tables that contains transactions worth millions to meet SEPA guidelines.
- Scheduled mainframe jobs on yearly, monthly, weekly, or daily basis using Control-M resulting in continuous data integration with SAP Concur to process and reimburse travel expense vouchers.
- Made use of Agile tools such as JIRA to document deliveries and test cases aiding towards better documentation.

Programmer Analyst| Cognizant Technology Solutions, India **Sep 2010 – Oct 2013**
Northern Trust Corporation

- Supported back office operations by providing reusable custom management reports using SQL,COBOL and Crystal Reports.

VIPIN DAS

Vancouver, Canada +12368632533 vdas@sfu.ca vipin-das-09875037
https://placemedas.github.io/ https://github.com/placemedas

ACADEMIC PROJECTS

Deviation Finder :An Elevator Anomaly Detection System (Github Link | Video Link) **Jan – Apr'20**

- Developed a system to efficiently predict anomalies and identify elevators that are at risk of breakdown.
- Conducted data pre-processing activities using Spark to process 75GB of accelerometer data spanned across 15 lifts and performed exploratory data analysis using Pandas, Matplotlib and Tableau.
- Applied feature engineering packages such as tsfresh to generate new features and experimented various machine learning models : Standard Anomaly detection, K-means, ANN Autoencoder and LSTM Autoencoder using Spark ML, TensorFlow and Scikit Learn.
- Evaluated the models using confusion matrix and implemented the model with the highest F1 Score(LSTM Autoencoder: F1 Score of 0.67).
- Created a scalable streaming data pipeline using Kafka, Tensorflow, Dynamo DB, Cron, Plotly Dash and Flask to serve the model.

Signature Verification using Convolutional Siamese Neural Networks **Jan – Mar'20**

- Modeled a Convolutional Siamese Network to verify and capture fraudulent signatures using state of the art method cited in *'SigNet: Convolutional Siamese Network for Writer Independent Offline Signature Verification'*. Pytorch was the tool used to perform this learning activity.
- A blog titled 'Do more with Less Data - One shot Learning with Siamese Neural Networks' is published in Medium.com

Crime Analysis and Prediction of crimes in major cities across Canada (Github Link) **Oct – Dec'19**

- Constructed a GBT Regressor model to predict crime rates in Canadian cities using Spark ML.
- Fetched crime data from various open data sources with different formats such as CSV,JSON using Python REST API calls to collate segregated data.
- Applied data aggregation, cleaning and structuring using PySpark to produce structured format.
- Modeled a NoSQL database using Cassandra DB to store scalable information.
- Created a dashboard of aggregated data using Tableau to generate insights.

Child Face Prediction using DCGAN(Github Link) **Oct – Dec'19**

- Modified a Deep convolutional GAN to generate the image of a child using the facial image of parents.
- Extracted the facial images of parents using multiple convolutional layers with the help of Pytorch.
- Experimented multiple approaches such as weight initialization and layer changes to improve the model and verified evaluation metrics such as SSIM and PSNR to quantify the similarity of the images.

EDUCATION

❖ **Simon Fraser University, Canada** **Sep'19 – Present**
Master's in Computer Science(Big Data Specialization) – CGPA 4.00
Courses - Machine Learning, Programming in Big Data, Programming in Data Science, Big Data Algorithms

❖ **Anna University, India** **Aug'06 – Apr'10**
Bachelor's in Computer Science and Engineering – 80%
Major Courses - Data Structures, Algorithms, Database Systems, Computer Programming, Data Mining

AWARDS

- Received **Best Project Award** from SFU Graduate Studies with a scholarship of CAD 3000 for creating an Elevator Anomaly Detection System.
- **'Excellence'** Award by Volvo for exhibiting successful project delivery with zero defect - Q4'2018.
- **'Outstanding Performer of the Year'** achievement for contributing towards cost saving initiatives - Mindtree ,2016.

SELF DIRECTED LEARNING

❖ GCP Essentials Badge, Google – Sep'20
❖ Programming with Python, Udemy - Aug'19
❖ Big Data Fundamentals, IBM - Aug'18
❖ Scientific Data Processing in GCP – Sep'20
❖ Elements of Data Science, AWS Training - Aug'20
❖ Data Science Fundamentals, IBM - Aug'18