

Vaibhav Sharma

Consultant

Dedicated and inquisitive professional with 5+ years experience in IT industry. Passionate about simplifying my work using script automations and keen to use my analytics solution-building exposure in enabling business transformation and processes. Learning to tell stories from data by leveraging my development experience to build solutions for leading clients. Having a good exposure with Python scripting and its data analytics suite, along with end to end delivery.

Experience

2019-06 - present	<div>Consultant</div> <div>Deloitte USI Consulting</div> <div>Mars Incorporated</div> <div>Achievements and Tasks</div> <ul style="list-style-type: none">Developed base calibration utility on for performing model selection across a range of 16 models , utilizing correlation , Z score and Percentage Changes in lieu with MAPE.Developed pipeline for FBProphet and Auto Arima based on historical data leveraging Data Bricks on Azure using PySparkDeveloped a parallel multiprocessing Base Calibration Segmentation utility to bucket and classify similarly performing model time series models on Data Bricks using PySpark.Participated in designing and user stories utilizing Jira and performed bug fixes on backlogs and new sprints following Scrum development practices <div>AmeriGas Partners LLP</div> <div>AmeriGas partners is the leading propane retail distributor in the United States. Assisted the client in understanding Customer Churn propensity and its leading causes.</div> <div>Achievements and Tasks</div> <ul style="list-style-type: none">Analysed and processed data from multiple sources for data modelling , using MS SQL , spanning across billion data points.Developing a flexible data transformations and pre-processing pipeline for model ready input.Developed a parallel multiprocessing Customer Segmentation utility to bucket and update customers based on there historical transactions.Developed and analysed multiple models namely , Logistic Regression ,XGBoost and LightGBM for customer churn propensity across multiple clusters.Developed a seminal execution pipeline for generating predictions as a module , increasing portability and reducing manual intervention. <div>MPI Analytics</div> <div>An internal product based on providing insights for pharmaceutical drugs in pre and post- launch sales across the entire country further granulised by HCP.</div> <div>Achievements and Tasks</div> <ul style="list-style-type: none">Worked as a Data Lead & managed a team of 2 members; helped in developing data model for assets & its supporting dashboards.Architected solution framework, keeping each asset loosely coupled, enabling flexible, robust & independent monolithic management resulting in 25% increased team velocity & before time deliverable.Assisted the development of Sales Forecasting asset , comprising of TimeSeries and LinearRegression based machine learning predictionsDeveloped standalone seminal wrappers for common interactions across the execution pipeline, enabling code reusability by 30% <div>Boehringer Ingelheim</div> <div>CoPay PCAT Analytics , Leveraging GCP & its cloud offering to accomplish a solution providing multiple business intelligence dashboard(s) which provides a bird eye's view of major products, including key metrics.</div> <div>Achievements and Tasks</div> <ul style="list-style-type: none">Spearheaded data ingestion and analytical framework from scratch to help support the deliverable dashboard, and intents related to major KPI's dynamics.Developed a python module based on bigquery-client with a dynamic and customisable rejection check, utilising the sweet spot between BigQuery and DataFlow, thus reducing the ingestion time by 10x folds & increasing raw data accuracy by 30% across 30 billion data rows.Lead the execution framework in seminal ingestion module(s) and actively coordinated the architectural design from offshore. <div>FedEx HR Analytics</div> <div>FedEx workdays HR analytics spans across regional employee data.Leveraging GCP Suites utilities to build pipelines and developed data-models around this data for driving better insights.</div> <div>Achievements and Tasks</div> <ul style="list-style-type: none">Worked as a Data science consultant , developing data pipelines for batch processing.Developed Incremental ingestion and ABC framework for SCD type - 1 & 2 changes leveraging DataFlow.
2019-01 - 2019-05	<div>Associate Data Scientist</div> <div>UnitedHealthGroup - Optum Global Services</div> <div>RADA</div> <div>RADA assists various stakeholders under the UHG umbrella to automate and provide insights related to their business segments. VIVO a leading project for Voice to Text Transcription , provides Analytics solutions tailored to each requirement and business segment.</div> <div>Achievements and Tasks</div> <ul style="list-style-type: none">Automated training Acoustic Models for CallTranscriptionEngine based out of HMM Kaldi toolkit , targeted for various business segments and languages, achieving ~18% WER (Word Error Rate).Undertook seminal development of CallTranscriptionEngine, utilising MultiProcessing thus reduced overall execution time by 60% & enabling multiple requests at once end to end.Developed a data preprocessing , feature extraction module for data cleaning and mining , targeted for unstructured transcribed data output for downstream workflows.
2015-07 - 2018-12	<div>Developer Analyst</div> <div>TCS</div> <div>Achievements and Tasks</div> <ul style="list-style-type: none">Worked across multiple Projects , primarily focused with application enhancement and development.Served as SPOC for multiple engagementsCreated multiple automation and wrappers to support core business operations enabling 25% increase in throughputLead and mentored, peers across multiple levels

Education

2010-07 - 2014-07	<div>Graduation</div> <div>RGPV University</div> <div>B.E (Electronics and Communication)</div> <div>CGPA - 6.9</div>
2009-05 - 2010-05	<div>HSC</div> <div>Bhavan's Prominent School, Indore</div> <div>Score - 69 %</div>
2008-05 - 2009-05	<div>SSC</div> <div>Bhavan's Prominent School, Indore</div> <div>Score - 69 %</div>

Additional Solution Experience

<div>Fuzzy Logic based Data DeDuplication</div> <div>Developed a Python module from scratch leveraging fuzzy logic algorithms to cluster similar records, aimed at eliminating duplication. The model clusters records into a single record based on edit distance and token set ratio vectors, creating a similarity matrix across a common hash space across multiple fields. The solution is aimed at Master Data Management, solution identified 37% duplicated records for a leading Pharma Client ranging across 330K records</div> <div>Technique used:- Leaders Clustering, Levenshtein & Edit Distance , Token Set Ratio</div> <div>Tools and Technology Used:- Python, Fuzzywuzzy, GCP</div> <div>Time Series based Anomaly Detection</div> <div>Built a time based model to identify anomalous data point(s). Input dataset was trained on lag values to predict & evaluate if generated trend values lie within the upper and lower boundaries , divergence between actual and predicted values are further evaluated upon a combination of static & sliding window by analyzing z-scores for classification. The model was evaluated across a benchmark Yahoo S5 labelled dataset.</div> <div>Technique used:- FBProphet, ARIMA, Auto ARIMA</div> <div>Model Validation:- Confusion matrix , Recall vs Precision , Z Score</div> <div>Tools Used:- Python (Pandas,Numpy,Sklearn,Scipy etc)</div> <div>PyModBus Concurrency Wrapper</div> <div>Developed a concurrency python script based on asyncio & Modbus protocol, to retrieve holding registers across multiple units values which were further ingested to MySQL DB using POST REST API. The script was scheduled using crontab and deployed on Raspberry PI</div> <div>Tool Used -: Python , PyModBus , Raspberry Pi, Linux</div> <div>Recommendation System</div> <div>Developed a recommender based system to predict movie recommendations based on two similarities - cosine and pearsonr. Model is further divided into Item - Item based and User-User based collaborative filtering. MovieLens dataset was used for the model with 862 unique customers with 1263 unique movies</div> <div>Technique Used -: Collaborative Filtering</div> <div>Tool Used -: Python (Pandas,Numpy,Sklearn,Scipy etc)</div>
--

Achievements

- Received Outstanding Performer & on the spot awards for various automations and achievement in multiple projects
- Coursera Certification on Data Engineering, Big Data and Machine Learning on GCP
- Microsoft - DAT210x (Python for Data Science) Certification from EDX
- Internal TCS certifications acquired for both Perl and Python till E2 level

Interests

- Table Tennis
- Gaming
- Driving
- Hiking

Personal Info

Address	C -156 Vikaspuri New Delhi
Phone	9752236647
E-mail	vaibhav.sharma12@hotmail.com
Date of birth	1992-08-12
LinkedIn	https://www.linkedin.com/in/vaibhav-sharma-280781136/
GitHub	https://github.com/vaebhav
StackOverflow	https://stackoverflow.com/users/9108912/vaibhav-sharma
LeetCode	https://leetcode.com/user4351t/

Skills

Python	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>
BigQuery	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>
C++	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>
SQL	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>
Google Cloud Platform , AWS	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>
NLP	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>
Machine Learning	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>
Internet of Things	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>
PySpark	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>
Data Warehousing	<div><div></div><div></div><div></div><div></div><div></div><div></div></div>