MANISH GUPTA

7760363070
manish9857@gmail.com

SUMMARY

- 7+ years of industry experience as a Data Scientist.
- M.Sc. (statistics) from Kalyani University in 2015.
- Experience in ML, Statistical Analysis, NLP, **Python**, **PySpark**, R, MySQL etc.
- Worked on different domains like Retail, Network Monitoring, Digital Marketing etc.
- Worked on algorithms like linear/log-linear regression, XGBoost, ARIMA, fbProphet etc.
- Successfully delivered project lead and as a part of team.

SKILLS

- Python
- PySpark
- SQL
- Azure
- R/ R Shiny
- PyTorch
- Regression (Linear, Log-linear, XG Boost, etc.)
- Classification (Logistic, Random Forest, Naïve Bayes, etc.)
- Time Series (ARIMA, ARIMAX, fbProphet, etc.)
- Cluster (K-Means, DB-Scan, Nearest Neighbors, etc.)
- Deep Learning (ANN, LSTM, CNN, etc.)
- NLP

EXPERIENCE

NOV'2022 – JUL 2023

ASSISTANT MANAGER, KPMG.

Project Undertaken

Objective: To monitor and maintain a retail forecasting model

- Monitor a Time series model running on Azure ML.
- Creating the checklist to flag any failure.
- Fix latency in model run and data pipeline.
- Helped set up QC checks, align with DE QA checks and set up scheduled reports for reliability.

JAN'2021 – NOV 2022

DATA SCIENTIST, TIGER ANALYTICS PVT LTD.

Project Undertaken

Objective: To create and automate SRM tool for a top FMCG company

- Set up and automated an ELT process using PySpark in Azure data bricks and Azure Data Factory.
- Created backend code for breaking down the product sales to different sales driver components.
- Created backend for Pre-Post analysis of promo events, highlighting performance of past and present promo events and predicted performance of promo planned for future based on historical data.
- Worked on backend for Promo efficiency view, highlighting effectiveness of different promo events.
- Worked as both individual contributor and team lead (DS).
- Owned and delivered several threads of the project as an individual contributor.
- Led a team of 4 DS resources, managing daily tasks and overall deliverables from Data Science part.
- Helped set up QC checks, align with DE QA checks and set up scheduled reports for reliability.

Objective: To analyzes and recommend digital campaigns on Google Ads and social media campaigns for a top FMCG company.

- Analyzed performances of different keywords bids from historical data to point out what is working and what is not.

- Worked closely with CRM to align with client focus area and provide them with correct information.
- Helped analyzing performance of different ad types on targeted customer segments and help create campaigning construct for better engagement on social media ads.
- Worked as an individual contributor and owned Google Ads thread.

NOV'2018 – OCT'2020

DATA SCIENTIST, THIRDEYE DATA ANALYTICS SERVICES INDIA LTD

Project Undertaken

Objective: To create an Auto-ML application to detect Anomalies in different source of data related to server activities.

- The data came from several sources (metrics). These sources related to activity logs like web-server activity logs, CPU utilization from different machines, local servers, etc.
- Created an ensemble of two approaches – threshold-based approach and Forecast approach. Used different approaches to create several layers of anomalies and used the rating method to flag points as anomaly or not.
- Threshold based approach included robust z-score to calculate upper and lower limits. Forecast approach included two algorithms: Prophet and LSTM. Used the time-series algorithms to predict future activities and detected anomaly by checking the deviation of actual from predicted value.
- Acted as a Senior Data Scientist, led the team with two more contributors, was responsible to come up whole approach and choosing algorithms, initial demo to client and day-to-day client handling.

Objective: To develop an algorithm to calculate catchment population for all the hospitals in a country.

- Created a modified version of Voronoi diagram using SciPy and geometry package and optimized it using google distance API to calculate catchment area for a network of hospital in a country, where the coordinates of the hospitals were provided by the client.
- Used worldpop.org to get population per pixel data for the country and mapped it with the catchment area of hospitals to figure out how much population is dependent on a hospital for health services.

DEC 2017 – TO OCT 2018

DATA ANALYST, THEOREM INDIA PVT LTD.

Project Undertaken

Objective: To classify a webpage to different categories based on its content.

- Used Google API and python to scrape thousands of web pages and extract relevant data.
- Used NLP to preprocess and Naïve Bayes to create classification model to classify a given webpage into different classes based on functionality
- Owned the project and worked individually form end to end.

DEC 2015 – TO JUN 2017

STATISTICIAN, SPSS SOUTH ASIA PVT LTD.

Project Undertaken

Objective: A petroleum conservation department wanted to improve the impact of their field activities.

- Explored the data to determine the performance of activities in different states.
- Used K-means cluster analysis to determine high performing and low performing activities.
- Used CHAID algorithm to determine the factors affecting the performance of the activities.

EDUCATION

- M.Sc. in Statistics from Kalyani University in 2015
- B.Sc. in Statistics from Calcutta University in 2013
- Higher Secondary in Science from Central Model School Bkp in 2009
- Class 10 from Central Model School Bkp in 2007