Surendra Kumar

Data Scientist | A Google Cloud Certified - Professional Data Engineer | Problem Solver

Ahmedabad | +91 8610292768 | sur365@gmail.com
linkedin.com/in/surendrakumar365/   skumarlabs.github.io

SUMMARY

Data Scientist & a Google Cloud Certified - Professional Data Engineer with total 4+ years experience in building scalable pipelines on GCP, an excellent understanding of NLP techniques, web scraping, building end-to-end predictive models and deploying them on cloud platforms like GCP, AWS etc

WORK EXPERIENCE

Associate Consultant   Jan 2020 - Current
Servian
At Servian, as an associate consultant, my role include design and deliver scalable data analytics & cloud solutions to the customer that help them sustain competitive advantage

Data Scientist   Jan 2018 - Dec 2019
Tatvic Analytics
Working at Tatvic, my domain includes predictive analytics focused towards applications of machine learning to increase the user engagement with the client's product and to develop machine learning powered enterprise solutions on Google Cloud Platform.

System Engineer   Oct 2015 - Dec 2017
Tata Consultancy Services Limited
Working as System Engineer in BFS domain, my role involved developing and maintaining solutions to gather, analyse client's data and generating insights and reports to help them better understand their customer.

PredictN For Remarketing

Description: A Prediction SaaS platform which predicts qualified audience list for remarketing who carry the higher probability of conversion on your site within the next 'N' days

Responsibilities:
- Extracting data from Google Analytics on GCP VM
- Preparing data using Google Cloud Datalab and DataPrep
- Training Machine Learning models to generate the optimal list of audience using Cloud ML and open source libraries like TensorFlow
- Deploying the model on Google Compute Engine for serving
- Creating performance report on Google Data Studio

Achievements:
- Improved CPL of insurance company AegonLife by >55% using this solution
- Improved CVR of Eureka Forbes by 30%. Details here
- Similar improvements were measured for one of the largest vehicle & health insurance company in India which used the solution to target only a specific segment of the audience to remarket instead of traditional remarketing approaches

Khas Apke Liye

Description: A content-based recommendation engine to serve better & personalized news articles on the website of *Jagran* Hindi daily having 31M+ readers per month

Responsibilities:
- Extracting newly published articles using CMS API and processing them using DataFlow pipeline
- Training a LDA content model on the corpus of Hindi articles & building user preferences using GCP VMs and BigQuery
- Deploying the model on Google Cloud Functions for serving the recommendations as JSON

Proficiency

Workflow
Visual Studio Code, Git, Docker. Remote Development on Linux

Machine Learning Libraries / Tools
TensorFlow, Keras, NumPy, Pandas, Scikit-Learn, Catboost, Matplotlib, Seaborn, Google Analytics, PySpark, NLTK, Gensim, Stanford CoreNLP, RapidMiner

Python 3
Modular and efficient code writing abilty with really good debugging skills

Google Cloud Platform
Compute Engine, App Engine, Kubernetes Engine, Cloud Functions, Datastore, Cloud Storage, Networking, Stackdriver, Dataflow, BigQuery, ML Engine, Cloud Migration

Database and Warehouse
BigQuery, DataStore, MySQL, MongoDB

Web Development
HTML, JavaScript, CSS, Flask, FastAPI

Core Competencies

- Predictive Modeling
- Deep Learning
- Statistics
- Time Series Analysis
- Topic Modeling
- Web Scraping
- Dialogflow
- Google Cloud Platform
- API development
- Google Analytics
- Data Studio Reporting

Languages Known

- English (Proficient)
- Hindi (Proficient)
- Kannada (Read & Write)
- Tamil (Read & Write)

Achievements:
- Achieved >8.9% CTR with the highest being >15.0% on mobile devices
- Improved AvgSessionDuration by >5 minutes

Social Listening

Description: This project involved social media analysis for a fashion brand and an online shopping giant to give insights into how their brand was perceived by their users

Responsibilities:
- Extracting data using Twitter API and Play Store reviews scraping
- Processing large corpus using distributed DataFlow pipeline
- Building performance metrics such as brand engagement with the users on social media, performing sentiment analysis and topic modeling on data
- Preparing reports & dashboards for insights in Google Data Studio

Other Side Projects

TV Ad Analysis: Time-Series analysis using RNNs to assess the impact of ad spots on target demographics & provide recommendations for ad spots yielding max impact

Feature Attribution Analysis for Conversion: Feature attribution analysis using Markov chain for our client's flights booking data involving analysis of the website features; then attributing success (Conversions & Revenue) to different features. Provided useful insights to the client like poor performing features, how to optimize UI etc

EDUCATION

Bachelor of Technology in Computer Science & Engineering   2011 - 2015
Harcourt Butler Technical University, Kanpur

An aggregate of 67.5%. Got exposed to various ethical hacking related activities and worked as lead in the technical team of the department as well.

Senior Secondary School Examination   2009 - 2010
Jawahar Navodaya Vidyalaya, Kanpur

An aggregate percentage of 80.2 in CBSE Board

Secondary School Examination   2007 - 2008
Jawahar Navodaya Vidyalaya, Kanpur

An aggregate percentage of 89.2 in CBSE Board

HACKATHON PARTICIPATION

- 12th rank in MakeMyTrip Data Science Challenge
- 25th rank in Caavo Vision Challenge
- 47th rank in Data Science Challenge on HackerEarth which involved classification of X-ray images
- 180th rank with AUC score 0.8478 on McKinsey Analytics Online Hackathon - Healthcare Analytics

CERTIFICATIONS & COURSES

- Google Cloud Certified Professional Data Engineer
- Networking in Google Cloud Platform training at Google, Gurgaon on May 7, 2019 to May 8, 2019
- Machine Learning with Tensorflow on Google Cloud Platform by Google on Coursera
- Deep Learning A-Z: Hands-on Artificial Neural Networks on Udemy Inc
- Machine Learning A-Z: Hands-On Python & R In Data Science on Udemy Inc
- Building Spark applications on Safari Online