






SURENDRA KUMAR

Kanpur Nagar, UP, India 
+91 8610-292-768 
sur365@gmail.com 
</in/surendra.kumar365> 
skumarlabs.github.io 



OVERVIEW

A Google Cloud Certified - Professional with more than 5 years of experience in building scalable pipelines, solutions 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.



SKILLS

- Python, C++, Java, SQL, YAML
- Spark, Kafka, Airflow, MongoDB, Neo4J
- **Google Cloud Platform**, Bigquery, Pub/Sub, AI Platform, Kubernetes, Dataflow, Compute Engine, Cloud Storage, Fire store, Data Fusion, Networking, Stack Driver, ML Engine
- **Text Processing**, Topic Modeling, Dialogflow, Web scraping, Text Classification, NLTK
- **Web Development**, Flask, Django, MySQL, Git, API Development
- **Machine Learning**, Time Series, TensorFlow, Keras, scikit-learn, Catboost, Random Forest, XGBoost
- **Statistics**, NumPy, Pandas, Seaborn
- Google Analytics, Data Studio Reporting
- Linux, VS Code, Jupyter Lab, Notebooks, Docker



EXPERIENCE

Sr Data Engineer | MediaAgility

05/2020 - PRESENT

Helping our clients to migrate to GCP & realize its full potential by proposing optimal solutions. Manipulating, processing, and extracting value from datasets using Big Data technologies and solutions

Data Scientist | Tatvic Analytics

01/2017 – 12/2019

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 | Tata Consultancy Services

10/2015 – 12/2017

Developing and maintaining solutions to gather, analyze client's data and generating insights and reports to help them better understand their customer



PROJECT

Google Cloud Projects/Architecture Designs

Worked on several proof of concepts projects to demonstrate capabilities of Google Cloud to our new clients, proposed architecture designs for workload migration from on-premise/other cloud services to Google Cloud etc.

- Successfully migrated client's data pipelines from Cron based setup to Apache Airflow on Google Cloud
- Created a secure, isolated data room for client on GCP where the Client could bring their confidential data for analysis using Python/R data science environment on AI Notebooks and Bigquery
- A web application to retrieve data from various documents like Adhaar card, PAN card, Passport etc. using Document AI and store in a NoSQL database

Predictive Remarketing

A SaaS platform which predicts the qualified audience list for remarketing campaigns carrying higher probability of conversion within the next N days

- Extracted Google Analytics data on GCP VM and prepared data using Google Cloud Datalab and DataPrep
- Trained Machine Learning models to generate the optimal list of audience using Cloud ML and open source libraries like TensorFlow
- Creating performance report on Google Data Studio
- Improved CPL of insurance company AegonLife by >55%
- Improved CVR of Eureka Forbes by 30%

Personalized Recommendation System

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

- Extracting newly published articles using CMS API and processing them using Dataflow pipeline
- Training an LDA content model on the corpus of Hindi articles using gensim & building user preferences using GCP VMs and Bigquery
- Deploying the model on Google Cloud Functions for serving the recommendations as JSON
- Achieved >8.9% CTR with the highest being >15.0% on mobile devices
- Improved average session duration by >5 minutes

Other Projects

- Time-Series analysis using RNNs to assess the impact of ad spots on target demographics & provide recommendations for ad spots yielding max impact
- Social media impact analysis for a fashion brand and an online shopping giant to give insights into how their brand was perceived by their users during the marketing campaign run
- End to end website journey of users, feature attribution using Google Analytics export data in Bigquery
- A Dialogflow web application to dynamically ask survey questions to the end user



EDUCATION

B. Tech | Harcourt Butler Technical University

2011 – 2015

Computer Science & Engineering, Bachelor's degree

Senior Secondary School Examination | Jawahar Navodaya Vidyalaya, CBSE

2009 – 2010

Physics, Chemistry, Mathematics, Computer Science

Secondary School Examination | Jawahar Navodaya Vidyalaya, CBSE

2007 – 2008

Science, English, Hindi, Mathematics



CERTIFICATIONS & COURSES

- Google Cloud Certified **Professional Data Engineer & Professional Cloud Architect**
- *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**



HACKATHON PARTICIPATION

- 12th rank in **MakeMyTrip** Data Science Challenge
- 25th rank in **Caavo** Vision Challenge
- 47th rank in Data Science Challenge on **HackerEarth** involving classification of X-ray images
- 180th rank with AUC score 0.85 on **McKinsey Analytics Online Hackathon** - Healthcare Analytics



LANGUAGES KNOWN

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

HOBBIES

- 3D Modelling (Blender)
- Animation
- Piano