

# Sankalp Tomar

Sr. Data and Applied Scientist @ Microsoft

Email : [sankalptomar@gmail.com](mailto:sankalptomar@gmail.com)

Mobile : +919600113437

Portfolio : <https://www.sankalptomar.site>

## About me

Enthusiastic Full Stack Data Scientist eager to contribute to team success through hard work, attention to detail and excellent organizational skills with over **11 years** of successful experience in Machine Learning, NLP and Big Data Engineering. I love working on projects which require applying mathematical models either for simulating phenomenon or for emulating the real world for machine learning. Recognized consistently for performance excellence and contributions to success in IT industry. Expert in dealing with real-world large data sets on Cloud and On-Premise environments.

## Education



### **Master Of Engineering – Computer Science (Data Science)**

BITS PILANI

08/2018 – 07/2020



### **Bachelor Of Engineering (Electronics & Telecommunication Engineering)**

Dr. D.Y.Patil Institute of Engineering and  
Technology, Pune

08/2008 – 07/2012

## Key Strengths

### **Strategic Product Development**

- Machine Learning and NLP based applications for the IT organization in a 2 Trillion Dollar enterprise of 320K employees worldwide
- ML based, customized suite of software tools for the Sales and Service organization ~5M USD annually in productivity improvements
- Enterprise Recommendation Engines, Search Engines and Personalization Engines ~ 2.75M USD annually in revenue and client retention
- Proven experience of working through the full product cycle from initial design to the final product
- Regular practice of estimating planned work, identifying dependencies, and participating in sprint planning and scrum ceremonies
- Ability to communicate effectively product architectures, design proposals and negotiate options at senior management levels

### **Data Science and Machine Learning**

- Presented research paper in MLADS on Semantic search
- Generative AI for enterprise applications like Question Answer systems using LLM, RAG(Retrieval-Augmented Generation), Llamaindex, Semantic Kernel and Langchain
- Semantic Search Engines using ML and NLP techniques for enterprise applications using Transformers
- Recommender Systems for retail company home pages and mobile applications
- Predictive models for estimating customer retention and churn
- Feature engineering and productizing both offline and real-time machine learning pipelines

## **Leadership and Project Management**

- Leading a global, cross functional team developing NLP based enterprise products
- Set up from scratch and managed team with 5+ Data Scientists and Engineers
- Experienced in agile, scrum, waterfall based software development life cycle and PLM
- 9+ years in client facing technical roles
- A coach for new Data Scientists and Engineers

## **Technology and Tool Development**

- Statistical modeling for data driven decision making
- Extremely comfortable with big data analysis, machine learning, NLP and LLM
- Python, PyTorch and Tensorflow based ML application development
- Experience in Model Based Design with C# with target deployment on Data Centers
- Familiar with application and enterprise level security for connected devices and products
- Experience working with functions like Marketing, Operations, Regulatory bodies, Legal, and Finance
- Experience with different cloud based solutions – AWS, GCP, Azure , Docker

## **Technical Skills**

**Machine Learning :** Regression Techniques, Decision Trees, Random Forests, Neural Nets, Clustering Techniques, Predictive modelling, Statistical Analysis, Classification analysis, Reinforcement Learning, Ranking and Recommendation, Time Series Analysis.

**Generative AI :** LLM, GPT, RAG, Langchain, Llamaindex

**Natural Language Processing :** NLTK, BERT, Transformers, Deep Learning, RNN,DNN,CNN, LSTM

**Framework :** Django, Flask, PyTorch, Tensorflow

**Bigdata Processing & Scripting :** Hive, Spark, EMR, Hadoop, HDFS

**Cloud/On Prem Platforms :** AWS, Google Cloud Platform, AZURE, Databricks

**Database:** Hbase, DB2, Redshift, SQL DW, ADW, Oracle, PostgreSQL

**Programming Language:** Python, C#, Linux, SQL, Scala

**Libraries :** Python with Pandas, SciKit Learn, PySpark with SparkML, Keras, Spacy

**Workflow Scheduling :** ADF, Azure Devops – VisualStudio, AWS Lambda, ECS, Docker,Airflow, Kubeflow, Mlflow

## Work Experience



**Microsoft Corporation** is an American multinational technology company headquartered in Redmond, Washington and develops, licenses, and sells computer software, consumer electronics, and personal computers. Microsoft is best known for its Windows operating system and **Office productivity suite**, but the company also produces a wide range of other software products, including cloud services and gaming consoles. With a market value of over \$2 trillion, Microsoft is one of the world's largest and most valuable companies.

### Senior Data and Applied Scientist - May '22 to Date

**Office** is the **M365** organization at Microsoft.

- I am a key stakeholder for the Data Science, Engineering and Product teams of **Powerpoint and Graphics**.
- Built High Scale Content Management Platform which serves **1 Billion** content across multiple Microsoft products including Word, Excel, PowerPoint, Teams, SharePoint.
- Set up guidelines and best practices for deploying **Machine Learning** models in production on the cloud and to integrate Python based model with C# based applications.
- Developed and implemented **Lean Product Development** methodologies.
- Led the concept development and delivery of a **Semantic Search Engine** for Content service to access Image base content. This product improved on the relevancy of the search results by **85%**.
- Led the innovation and product roadmap for the delivery of a new solution that uses state-of-the-art **NLP LLM** using FLAN-T5 models. Using Indexing and Embedding techniques brought down the response time by **80%** and reduced space requirement by **97%** to store the embeddings and index.
- Saved **800K USD** annually by using multilingual model which infers the language and doesn't require any manual translation of the image description based on the locales.
- Built Transformers based Computer Vision models to recommend images and design ideas based on content of Slides which led in increase of Stock content use by **95%**.
- Integrating Dall-E, ChatGPT and Stable diffusion to provide Image editing and enhancement capability to users in Office wide products.
- Interviewed, Hired and set up a specialized science and engineering team with Applied Scientists and Data Scientists.



**Gartner, Inc.** (NYSE: IT), is the world's leading research and advisory company and a member of the S&P 500. **Gartner** equips business leaders with indispensable insights, advice and tools to achieve their mission-critical priorities and build immensely successful organizations of tomorrow.

### Lead Data Scientist - May '19 to May '22

Gartner is an Enterprise Research and Advisory company. Gartner's core offerings are Enterprise Market Research, Insights and Recommendations to CXOs such as **Magic Quadrants** and **Hype Cycles**. I led a cross functional team of geographically distributed data scientists, data engineers and software engineers for web and SaaS application development. My product portfolio consisted of **2 applications which serve 20,000 premium clients (CxO Suite) and 6000 internal associates**.

- Led a team to develop cloud based, decision support systems that enable the Service teams for customer retention.
- Designed and developed **Natural Language Processing (NLP)** based applications to extract insights like market intelligence from public data sources such as 10K documents, Earnings Call Transcripts, SEC filings, News, etc.
- Designed a **Recommendation System** to provide Gartner users with a highly personalized feed of recommended items on their home screen and on the mobile app.

- The algorithms powering this product recommend a rich set of Gartner's offerings like research documents, Peer Connect conversations, Events and Webinars in a highly personalized format to our clients that translated into an approximate dollar amount of **1.25 Million USD annually**.
- Developed predictive algorithms for estimating customer risk and churn.
- Stipulated the workflows for setting up data pipelines for the Machine Learning models in production.
- Developed NLP techniques to extract Key Questions, Topics from call transcripts and recommend research documents and SME's to better prepare the sales team.
- Increased customer retention from 42% to 39% which translates to about **1.5 Million USD per annum**.
- Developed hiring roadmaps and staffing requirements for projected product development plans.
- Coached and mentored new Data Scientists, Data Engineers and Software Engineers in the team.



**Publicis Sapient** is a global consulting company that helps businesses drive digital transformation through strategy, design, and engineering. With a focus on delivering innovative and customer-centric solutions, **Publicis Sapient** serves a diverse range of industries, including financial services, healthcare, and retail. Its services include digital strategy, experience design, software engineering, and data analytics, among others. Publicis Sapient is part of the Publicis Groupe, one of the world's largest advertising and communications groups.

## Senior Associate - Data Science - Nov'17 to Apr'19

**Client** : Mercedes Benz (DAIMLER) – OneDNA Platform

- Lead the team of 5 Junior Data Science Engineer to build end to end Machine Learning platform to get the insight of customer behavior which translated into increase of customer retention by **72%**.
- Built Machine Learning & Deep Learning models using 100+ GB's of Adobe clickstream data.
- Created spark-based workflows for continuous ingestion of Data and perform Data processing and feature Engineering.
- Scaling the solutions for all 35 markets (different geos) using distributed environment.
- Integrating Model Output with Tableau dashboard which is primarily reviewed by **CxO Suite** of Mercedes Benz on regular basis.
- Deployed the model on Azure environment using Azure Data Factory and Databricks which runs in batch and on Ad-hoc basis.
- Demonstrating insights, models, mathematical and statistical concepts to non-technical stakeholders which helps them to plan the business strategies.
- Increased the **Test Drive booking by 40%** by addressing the key pain areas faced by the customer while booking test drive which translated to **increase in sales by 12%**, increasing the overall revenue of the organization by **5%**.

**Client**: Michael Kors

- Building Recommendation Engine based on the Item-Item similarity to recommend items based on the customer journey and item selected.
- Predicting the customers who are likely to be converted using Random Forest and XGBoost and assign them buckets for sales team to run campaign using email.
- Customer retention and conversion increased by **60%** translated into Increased the sales by **1.2 Million USD per annum** by better recommendation.
- Preparing data flow pipeline from FTP server using python and AWS.



## Sr. Software Engineer - Machine Learning - Sep'13 to Nov'17

Infosys is a multinational IT consulting and services company headquartered in Bangalore, India. Founded in 1981, Infosys provides a range of services including **software development, maintenance, and engineering** as well as consulting and digital transformation services and has a global footprint spanning over **50 countries**.

### Client: American Express

- Built PySpark based pipeline for data processing and wrote Hive Queries to fetch the data from CornerStone.
- Built K-Means clustering algorithms to detect and avoid fraud transactions which translated to decrease in fraud transaction by **20%**.
- Developed the Python code in Spark environment to generate the insights for the customers to help them make business decisions.
- Experienced with the Spark improving the performance and optimization of the algorithms in Hadoop using Spark Context, Spark-SQL, Data Frame, and Pair RDD's.
- Responsible for design & development of Spark SQL Scripts using Python.

### Freelance Experience

- Worked with a Startup which primarily works in Computer Vision and NLP Field as a freelancer.
- Built a product from scratch to :
  - Classify the documents in PDF or JPG format to different classes (i.e. Invoice, Resume).
  - Extract NER from the documents.
- **Technology used:** PyTorch, Flask, AWS, CNN, Neural Networks, Open CV

### Certifications

- Deep Learning and Machine Learning with Python from IBM.
- Machine Learning Specialist from LinkedIn
- Google Certified Professional Data Engineer