

**Sumit Kumar**

**E-mail:** [sumit.datascientist@gmail.com](mailto:sumit.datascientist@gmail.com)

**Mobile :** +91 9873788393

<https://sumitdatascientist.github.io/>

<https://www.linkedin.com/in/sumit-datascientist/>



---

---

### Profile Objective:

To be associated with a progressive organization, where I can employ **Data Engineer, Associate Machine Learning Engineer** knowledge and skills to contribute the growth of the organization and personal growth in an effectual manner.

- ⌘ Offering **2.5+ Years** of experience can be headhunted for a Lead level position across any functional sectors within an IT organization of repute
- ⌘ Solid programming skills with **Python and its libraries** like **NumPy, Pandas, Matplotlib, Seaborn, Scikit-Learn**, etc.
- ⌘ Experience in Developing Spark applications using **Azure Spark - SQL in Databricks** for data extraction, transformation and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns
- ⌘ Good understanding of **Azure Spark Architecture** including **Spark Core, Spark SQL, Data Frames, Spark Streaming, Driver Node, Worker Node, Stages, Executors and Tasks**
- ⌘ Excellent communication skills with excellent work ethics and a proactive team player with a positive attitude
- ⌘ Domain knowledge of **Retails, Telecoms IVR, Online Food Ordering, Design, Logistics, Travel and Health**
- ⌘ Strong skills in visualization tools **Power BI, Confidential Excel - formulas, Pivot Tables, Charts and DAX Commands**
- ⌘ Expertise in various phases of project life cycles (**Design, Analysis, Implementation and testing**)
- ⌘ **Solid understanding** of **mathematical** foundations behind **Machine Learning Algorithms** and comfortable with discussing them in details.

### Technical Skills

- ⌘ **Programming Languages:** Python, Spark Python, Spark SQL, SQL, Scala
- ⌘ **Python Libraries:** Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn
- ⌘ **Azure Architecture:** Data Lake Storage, Data Factory, Databricks. SQL Database
- ⌘ **Databases:** SQL Server 2014/2012/2008/2005, MySQL, PostgreSQL
- ⌘ **Cloud Platform:** Azure
- ⌘ **Other Tools:** MS Excel (VBA, Macro), Power BI(DAX), Jupyter Notebook, PyCharm, Visual Studio
- ⌘ **Other Skills:** Machine Learning Algorithms, Basics in Deep Learning, Data Structure, Algorithm, OOPs

### Academic Details

- ⌘ Post Graduate Diploma in Data Science, Symbiosis, Maharashtra [2020 - Appearing].
- ⌘ Bachelor of Technology (ME) from IEC College of Engineering and Technology, (AKTU University) with **69.80%** in 2017.
- ⌘ HSC from BSEB Board with **74.60 %** in 2013.
- ⌘ SSC from BSEB Board with **76.40 %** in 2010.

## Professional Experience

**Webhelp [Jan 2021 – Present]**

**DLF Centre Court, Gurgaon**

**Associate Data Engineer**

### Responsibilities:

- ℵ Analyze, design and build Modern data solutions using **Azure Service** to support visualization of data. Understand current Production state of application and determine the impact of new implementation on existing business processes.
- ℵ **Extract Transform and Load** data from Sources Systems to **Azure Data Storage** services using a combination of **Azure Data Factory, Spark SQL and Azure Databricks. Data Ingestion** to one or more Azure Services - (**Azure Data Lake, Azure Storage, Azure SQL**) and processing the data in **Azure Databricks**.
- ℵ Created Pipelines for multiple clients in **ADF** using **Linked Services/Datasets/Pipeline/** to Extract, Transform and load data from different sources like **Azure SQL, Blob storage, Azure SQL Data warehouse**, write-back tool and backwards.
- ℵ Developed **Logical KPIS** as per client requirement using **PySpark and Spark-SQL** for data extraction, transformation and aggregation from multiple file formats for analyzing & transforming the data to uncover insights into the customer usage patterns.
- ℵ Responsible for estimating the **cluster size, monitoring and troubleshooting of the Spark Databricks cluster**.
- ℵ Developed **JSON Scripts** for deploying the Pipeline in **Azure Data Factory (ADF)** that process the data using the SQL Activity.
- ℵ Created Build and Release for multiple projects (modules) in **production environment using Azure Services**.
- ℵ Hands-on experience on developing **SQL Scripts** for automation purpose.

**Extension Technology [Aug 2019 – Sep 2020]**

**Faridabad, Haryana**

**Junior Software Developer**

### Responsibilities:

- ℵ Working on **SQL Queries** for **Data Visualization** and **Creation Dashboard**.
- ℵ Made a different type of **reports** in **Zoho Analyst** as clients needed.
- ℵ Got Recommended to Many Projects from Senior Managers.
- ℵ Worked on Zoho tools like Creator, Crm, Books, Inventory backend in **Python and Deluge Script**.
- ℵ API Integrated 3rd party applications like **India Mart, Flipkart, Websites** etc. to Zoho tools.
- ℵ Also Worked on **ERP Next Software Backend** and **ERP Next Implementation**.

## Certificate

Python 101 for Data Science: **Cognitiveclass** [6 Weeks]

Python for Data Science: **IBM** [ 4 Weeks]

5 Days Data Science Workshop: **GreyAtom** [ 5 Days]

Machine Learning: **Udemy** [4 Weeks]

Deep Learning: **Udemy** [4 Weeks]

## Endorsement

I hereby declare that the information furnished above is complete and true to the best of my knowledge.

**Place: India**

**Sumit Kumar**