# Ankit Baviskar

plot no 39 , sadhashiv nagar,Jalgaon,Maharashtra 425001
91-7219428546 ankitbaviskar3@gmail.com linkedin.com/in/ankit-baviskar b0a2051a4 github.com/ankitjb26

## Education

**IACSD CDAC** **Mar 2022 – Sep 2022**
*PG-Diploma in Big Data Analytics* *Pune, Maharashtra*

**SSBT COET** **Aug. 2017 – Sep 2021**
*Bachelor of Computer Science Engineer* *Jalgaon, Maharashtra*

## Experience

**New Vision Software** **Apr 2023 – Present**
*Data Engineer* *Pune, Maharashtra*

- Developed Spark applications using Scala for data processing and analysis.
- Developed and maintained data pipelines using Apache Spark and Scala
- Developed and maintained data pipelines using Azure Data Factory, enabling efficient data processing and transformation
- Orchestrated end-to-end data workflows, integrating data from various sources into target destinations.
- Managed and optimized data storage solutions including Blob Storage, Azure Data Lake Storage (ADLS), and S3 buckets.
- Created and optimized ETL (Extract, Transform, Load) processes to ensure the efficient flow of data from various sources to data storage.
- Managed and optimized data storage solutions including Blob Storage, Azure Data Lake Storage (ADLS), and S3 buckets.
- Ensured data accuracy and integrity through data validation and quality checks.
- Collaborated with cross-functional teams to understand data requirements and deliver effective solutions.
- Currently gaining basic exposure to PySpark and Databricks, enhancing skills in data processing and analysis.

## Projects

**Pothole and Garbage Detection Using Classification Algorithm** | *Python, Machine Learning,*
* In this project a model for the detection of garbage and potholes based on artificial vision and deep learning is proposed, which detects garbage and potholes present in the streets of a city.
* Implemented Support Vector Machine and RESNET50 Algorithm to classify Potholes and Garbage.

**Data Pipeline Development: Gathering, Transforming, and Warehousing Data** | *Azure Fabric ,pyspark,SSMS*
* Successfully gathered data from multiple sources including JSON, CSV, and real-time streams.
* Implemented robust data ingestion pipelines to ensure reliable and timely data acquisition.
* Designed and implemented a lakehouse architecture to store and manage the collected data effectively.
* Developed and implemented data transformation workflows to cleanse, normalize, and enrich the gathered data.
* Utilized tools and frameworks for data processing, such as Apache Spark, to handle large-scale data transformations efficiently.
* Established data warehouses for storing processed and transformed data, ensuring accessibility and query performance.
* Implemented incremental data capture mechanisms to efficiently identify and process new or updated data.

**Global Superstore Dashboard Using PowerBI.** | *PowerBI*
* Dashboard is made to find meaningful insights from online orders of a superstore in the US.From this data, we will analyse the sales and profits of the US superstores based on the discount rates of the products, city or states where people shop the most, customer loyalty and product discount rates, etc.

## Technical Skills

**Languages**: Python, Scala, Spark , Pyspark , SQL
**Developer Tools**: VS Code, IntelliJ Idea, MS Azure (ADLS , ADF,DATABRICKS,Fabric , AWS (S3 bucket , EC2 instance) , SSMS
**Technologies/Frameworks**: Linux, Cloudera

## Extracurricular

* Strong problem-solving and analytical skills.
* Excellent communication and teamwork abilities.
* participated in Hackathon Competition