Project Title: AI – Powered Data Validation and Standardization in Supply Chain

1. Abstract

This project proposes an AI-driven approach to enhance data quality and consistency within supply chain operations. By automating data validation and standardization processes, we aim to reduce manual effort, minimize errors, and improve decision-making. Supply chain operations depends heavily on accurate and consistent data to function optimally. However, data quality issues such as missing values, inconsistencies, and errors can significantly hinder efficiency and decision-making. It addresses these challenges by leveraging advanced AI techniques to automate data validation and standardization processes by continuously monitoring data quality metrics to proactively identify and address emerging issues.

2. Problem definition

Supply chain operations depends on a vast amount of data from various sources, including suppliers, manufacturers, distributors, and customers. However, the quality and consistency of this data often suffer from data inconsistency, missing data and data Quality Issues. These issues result in delayed decision-making, increased costs, reduced customer satisfaction and compliance risks.

Key Questions:

- How to enhance supply chain resilience, mitigate procurement risks, and optimize supplier management strategies to meet market demand?
- How will you assess the quality of your existing supply chain data?
- What techniques will you use to clean and preprocess data, such as handling missing values and inconsistencies?

Target users:

- **Supply Chain Analysts and Managers:** Responsible for monitoring and analyzing supply chain performance. Need accurate and timely data to make informed decisions.
- **Data Analysts and Scientists:** Involved in data cleaning, transformation, and analysis.

Goal:

- To reduce errors and inconsistencies in supply chain data and enhance data accuracy, completeness, and timeliness.
- To provide reliable and accurate data for informed decision-making.

3. Requirements

Functional Requirements

- Transform and clean data to a standardized format.
- Validate data in real time to prevent error from propagating through supply chain.
- Train and retrain the machine learning models to improve accuracy and performance overtime.

Non-Functional Requirements

- Able to handle increasing data volumes and complexity.
- User-friendly interface, easy to understand.

4. Tools and Platform

4.1 Tools

- **Data Preprocessing:** Python, Pandas, NumPy
- **Model Development**: Scikit-learn
- **Visualization:** Power BI, Matplotlib or Seaborn

4.2 Amazon Web Services (AWS)/ IBM Cloud Services

Amazon Web Services (AWS)

- **Data Storage:** AWS S3
- **Data Preprocessing:** AWS Glue/AWS Lambda
- **Model Training and Deployment:** Amazon SageMaker/AWS Lambda

IBM Cloud Services

- **Data Storage:** IBM Cloud Object Storage
- **Data Preprocessing:** Jupyter Notebook
- **Model Training and Deployment:** IBM Watson Studio

5. Literature Survey

- "Revolutionizing Supply Chain Management With AI: A Path to Efficiency and Sustainability": This paper addresses a critical gap by presenting a comprehensive framework that not only highlights the benefits but also explores the limitations and challenges of AI adoption in real-world supply chains.

- "Development and Implementation of a Standardized Supplier Certification and Validation Framework for Enhanced Pricing, Lead Time, and On-Time Delivery Performance": This research paper addresses the need for a data-driven framework that incorporates safety, reliability metrics, and technical competencies, alongside traditional performance indicators, to ensure comprehensive supplier validation and effectively address challenges and consumer complaints in supply chain management.

6. Implementation Plan

Step 1: Data Preparation

- Upload Cosmetics dataset to **Google Colab Notebook/IBM Cloud Object Storage.**
- Clean the data to remove duplicates and handle missing values.

Step 2: Model Development

- Train a simple isolation forest model and DBscan model using Scikit-learn in **IBM Watson Studio (Jupyter notebook).**

Step 3: Deployment

- Deploy the model in **IBM Watson Studio.**

Step 4: Reporting and visualization

- Visualizing is done in Power BI / Matplotlib or Seaborn.

7. Expected outcomes

- Improved Data Quality by Reducing Errors of the Significantly lower rates of data entry errors, inconsistencies, and inaccuracies across all supply chain data sources.
- Increased Efficiency and Productivity by automation of data validation and standardization tasks.