

# Biplov Singh

♥ Jamshedpur, Jharkhand | ☎ +91-8969267638 | ✉ biplovsingh785@gmail.com

🌐 LinkedIn | 🐙 GitHub

## PROFESSIONAL SUMMARY

---

Aspiring Data Engineer with experience in Python, SQL, ETL pipelines, and big data technologies. Skilled in building scalable data processing workflows, analyzing structured datasets, and developing data-driven applications using Hadoop ecosystem tools. Passionate about distributed systems, data infrastructure, and transforming raw data into actionable insights.

## EDUCATION

---

- **Masters in Data Science** 2025-27  
*Indian Institutes of Information Technology, Lucknow*
- **Bachelor of Science in Computer Application and Cyber Security** 2022-25  
*Jharkhand Raksha Shakti University, Ranchi* CGPA: 7.92

## PROJECTS

---

- **Clickstream Analytics Pipeline** Jan 2026 – Feb 2026  
*End-to-end big data pipeline for processing and analyzing website clickstream data using Hadoop ecosystem tools.*
  - Built a complete ETL pipeline using Apache Flume, Pig, and Hive to process website clickstream data.
  - Designed data ingestion workflows to collect and stream raw user interaction logs into the Hadoop ecosystem.
  - Containerized the pipeline environment using Docker for reproducible deployment and testing.
- **Automated Silver-to-Gold Layer Pipeline** Feb 2026 – Present  
*Distributed ETL pipeline using PySpark and Gemini for large-scale review analytics.*
  - Designed a Medallion Architecture (Bronze, Silver, Gold) pipeline to process and transform large-scale Amazon review datasets.
  - Built distributed data processing workflows using PySpark for cleaning, deduplication, filtering.
  - Loaded transformed gold-layer data into PostgreSQL for downstream analytics and querying.
  - Developed scalable ETL scripts for automated layer transitions from Bronze to Silver and Silver to Gold datasets.
- **Crop Yield Prediction System** Nov 2025 – Dec 2025  
*Machine learning model to predict optimal crop choices based on environmental factors.*
  - Developed a predictive model using Python, Pandas, and Scikit-Learn to analyze agricultural data including soil metrics, temperature, and rainfall.
  - Implemented classification algorithms to recommend the most suitable crops, enhancing data-driven agricultural decision-making.

## ACADEMIC EXPERIENCE

---

- **Teaching Assistant** Aug 2025 - Dec 2025  
*Indian Institutes of Information Technology, Lucknow* Lucknow, UP
  - Assisted professors in managing coursework, designing assignments, and delivering supplementary instruction for undergraduate computer science and data science courses.
  - Conducted weekly doubt-clearing and practical lab sessions, guiding students through complex programming concepts and debugging.
  - Evaluated and graded student projects, assignments, and examinations, ensuring fair and constructive feedback to support academic growth.

## TECHNICAL SKILLS AND INTERESTS

---

**Languages:** Python, SQL, C/C++, JavaScript

**Libraries:** Pandas, Numpy, PyTorch, Matplotlib, Scikit-Learn, NLTK, Git, Docker, Kubernetes

**Data Engineering & Big Data:** Hadoop, Apache Hive, Apache Pig, Apache Flume, Yarn, ETL Pipelines

**Operating System:** Linux/Ubuntu

**Cloud/Databases:** PostgreSQL, MySQL, MongoDB

**Relevant Coursework:** Data Structures & Algorithms, Operating Systems, Object Oriented Programming, Database Management System, Big Data Analysis.

**Areas of Interest:** Data Engineering, Big Data Analytics, Distributed Systems, Machine Learning, Cloud Computing.

**Soft Skills:** Problem Solving, Self-learning, Presentation, Adaptability

## POSITIONS OF RESPONSIBILITY

---

- **Tech Support Volunteer** College Event - JRSU, Ranchi Aug 2024
  - Provided real-time technical assistance for an event, maintained continuous system availability for all presentations.
  - Troubleshoot connectivity issues.