

Somdev Pahari



Highly competent in **Hadoop**, **Spark**, **Kafka** and **ML** and seeking a career in **Big Data Analytics** in an organization which recognizes and utilizes my true potential while nurturing analytical and technical skills.

Technical Skills

Core Technologies	Hadoop, Spark, Hive, Kafka, Flume, Sqoop
Databases	MySQL, Cassandra
Languages	Python (PySpark, NumPy, Pandas, scikit-learn), Scala, Java
BI Tools	Tableau
OS	Windows, Linux

Experience

Projects

Fraud Detection System (Ongoing Capstone Project of Big Data Engineer PGP)

Objective: In this digital age, a leading bank in the world wants to safeguard its digital money, systems and data. This project is about building an application which can detect fraudulent transactions and identify lost cards and fraud endpoint that serves their purpose.

- **Spark ML** job reads the data from **Cassandra** and stores a trained model inside a storage.
- Spark streaming job loads the model, consumes transactions data from **Kafka** and stores predictions inside Cassandra.
- Real time fraud transactions come inside **Tableau** dashboard.

Chicago Taxi Data Research (Big Data Engineering PGP)

Objective: This project is about facilitating a taxi booking application to send appropriate promotional details to target customers by grouping them based on peak hour, low activity, origin-destination pairs etc. and also forecasted the growth of market size which assist executives and leadership teams for their business growth.

- Configured **Hive** with **HDFS** and executed **Hive SQL** on **HUE** to get the required data set.
- Using Java API data moved in **MySQL** and NoSql **Cassandra** databases for future querying/analysis.
- Used **Tableau** to display the data as per the use case requirement on dashboards.

Sentiment Predict (Udemy : Machine Learning A-Z: Hands-On Python & R In Data Science Certification Project)

Objective: Based on demand from a reputed restaurant, this application was built to predict the sentiment of the customer's review which will help them to upgrading their services if necessary.

- Bag-Of-Words model together with **NumPy**, **Pandas**, **scikit-learn** libraries in **Python** were used for performing several steps related to ML pipeline to achieve the result.

JAN 2018 - CURRENT

Senior Research Scholar / NIT Rourkela, Rourkela

Field of Research: Molecular dynamics (MD) simulation studies on biomolecules.

- Worked on several **Machine Learning** related projects by building **ML pipeline** using **scikit-learn** which helped to understand several critical biological phenomenon like bio-membrane curvature, diffusivity etc.

Education

APRIL 2020 - CURRENT

Post Graduate Program in Big Data Engineering / edureka! (Collaborated with NIT Rourkela), Bengaluru

Gained in-depth knowledge on various Big data concepts and did hands on like **MapReduce** using **Java**, **Spark** with **Scala** and **Python**, Data warehousing management with **HUE**(**HIVE** and **HDFS**), real time data streaming using **Kafka** and **Flume**, various **sorting algorithms**, **MySQL** and **Cassandra** data bases, **Big data analytics**.

JUNE 2015 - JUNE 2017

M. Sc. Physics / Vidyasagar University, Midnapore

Studied M. Sc. In Physical Science with 62.5%.

Certification Details

Udemy

- Machine Learning A-Z: Hands-On Python & R In Data Science 
- TABLEAU 2018 ADVANCED: Master Tableau in Data Science 

Languages

- English, Hindi, Bengali, Odia

Interests



ML and AI



Photography



Badminton



Camping

Personal Information

Name : Somdev Pahari
Contact No : +91 9593606626
Email : somdevpahari1@gmail.com
Location : Kolkata, WB