




# SOUMYA KANTA DASH

## HADOOP AND MACHINE LEARNING DEVELOPER

 [soumya.sk@gmail.com](mailto:soumya.sk@gmail.com)

 <https://www.linkedin.com/in/soumyaskd>

 +91-9658587829

 [https://www.hackerrank.com/soumya\\_skd1](https://www.hackerrank.com/soumya_skd1)

## PROFESSIONAL SUMMARY

Machine learning developer with 2+ years of experience in service industry. Skilled with a demonstrated history of working with machine learning, statistics, problem solving, and programming. Seeking to boost my career in Data science and acquire advanced skills in the process.

## SKILLS AND ABILITIES

- Regression Algorithms:
  - Linear
  - Random Forest
  - Gradient Boost
  - Support Vector
  - K-Neighbors
- Shell-Scripting
- Efficient and quick Debugging
- Data Warehousing

## TECHNOLOGIES/LANGUAGES

- Languages:
  - Python
  - Scala
  - C++
- Machine Learning:
  - Scikit learn
  - NumPy
  - Pandas
  - Matplotlib
  - Seaborn
- Big-data:
  - Hadoop
  - Spark
  - MapReduce
  - Oozie
  - Sqoop
- Databases:
  - Hive
  - Hbase
  - Oracle
  - PostgreSQL

## WORK EXPERIENCE

**Tech Mahindra Ltd. | Bhubaneswar** *Sep-2016 to Present*  
GridSolutions BI Analytics

- Fulfilled all expected machine learning and visualization duties for the Energy division of **General Electric**.
- Created and presented models for sales distribution over specific geographical areas to stakeholders reducing inventory related decision time by 20%.
- Created machine learning tools that predicted sales figures for upcoming quarters using Random Forest regression.
- Conducted unit testing, debugging and performance tuning of codes/scripts/queries.

### IQP to Hadoop Migration

- Completed migration of over 250 jobs from Informatica to Hadoop ecosystem successfully leading a team of mere 3 members in a span of 3 months.
- Slashed overall load duration by 50% by optimizing spark parameters for each job while ensuring the continuation and enhancements of services.
- Designed spark code for data transformation requirements using Spark DF and Spark while observing coding standards.
- Sequenced and scheduled Spark jobs using Talend to handle bottlenecks and optimize performance as well as resource utilization.

## EDUCATION

**Veer Surendra Sai University of Technology**  
**Burla, Odisha, India** *July-2012 to May-2016*

Bachelors of Technology in Production Engineering  
CGPA: **8.7/10**

## LEADERSHIP/COMMUNITY SERVICE

- Lead the departments of brakes, aerodynamics, engine and transmission in the team of Supra representing our University.
- Trained a several batches of new joiners on technologies like SQL, PL-SQL, UNIX and Big data fundamentals and received great response from peers and Project Leaders.
- Received 'BRAVO' award for learning and implementing multiple technologies in a short span of time.
- Collaborated with UMEEDIEN to spread awareness on various forms of cancer and referring patients to specialized doctors.
- Participated in CSR activity for teach fundamentals to adolescent children from deficient families to procure jobs.