Sourajyoti Datta

+91-9902600550 ☎
datta.sourajyoti@gmail.com ⌧

T1, #19, Darshan Nivas, 9th Main, SG Palya, CV Raman Nagar, Bangalore - 560093

Computer Science & Engineering graduate with 4+ years of experience with a keen interest and excellence in analyzing business requirements, problem solving, design, development and solution optimization and seeking a challenging Data Engineer/Data Science Specialist position in the areas of Analytics and/or Data Management.

ACADEMIC QUALIFICATION

Qualification	School/University	Specialization	Year	DGPA
Bachelor of Technology	Haldia Institute of Technology, West Bengal University of Technology	Computer Science & Engineering	2008-2012	8.15
Higher Secondary Examination	BSF Secondary School, Central Board of Secondary Education	Computer Science	2008	89.00%
Secondary Examination	Stepping Stone Model School, Council For The Indian School Certificate Examinations	PCMCB	2006	89.80%

PROFESSIONAL EXPERIENCE

1. TECHNOLOGY CONSULTANT
Analytics Data Labs, DXC Technologies/Hewlett Packard Enterprise Services, Bangalore

(Nov 2016 – Present)

Analytics and consulting engagement with DXC/HPE's commercial clients across verticals such as Freight and Automobile.

- **Automobile After-Sales Analytics – Leading German Automotive Corporation**
  - Analytics was used to identify the price elasticity of parts, and hence, calculate the optimum price point for each spare part.
  - POC was done on R and Spotfire Visualizations
  - Developed Hive SQL scripts for transformation(s) on staging tables.
  - Developed Spark jobs for the Statistical and ML algorithms
  - Spotfire visualizations were ported to Tableau for better performance
  - **Technologies:** R, Spotfire, Hive, Spark, Tableau

- **Driver Distance Analytics – Leading American Freight Shipping Services Corporation**
  - Analytics was used to support regulatory compliance and reduce potential software costs.
  - An algorithm was devised to calculate the distances travelled each day by all drivers in the system, identifying those that exceed the regulatory requirement.
  - Developed Vertica SQL scripts to transform data on Vertica into ingestible format for R-UDF.
  - Developed generic R-UDF codes for the models.
  - **Technologies:** R, Vertica, Spotfire

- **Predictive Analytics – Leading American Freight Shipping Services Corporation**
  - Built a predictive model to forecast Pickup & Delivery and Line Haul capacity that would support planned investments in P&D and Line Haul optimization software.
  - Analyzed cost and shipment data to derive key drivers impacting shipment volume.
  - Automated analytic models was developed for each of the top shippers to predict volume at shipment level.
  - Developed Vertica SQL scripts to transform data on Vertica into ingestible format for R-UDF.
  - Developed generic R-UDF codes for the models.
  - **Technologies:** R, Vertica, Spotfire

2. DATA SCIENCE SPECIALIST
Analytics Data Labs, Hewlett Packard Enterprise/Hewlett Packard, Bangalore   **(Jan 2015 – Oct 2016)**

Analytics and consulting engagement with HP/HPE's commercial clients across verticals such as Telecom, Transportation and POCs for Scalable Analytics COE.

- **Analytics Engine – Telecom Product**
  - o Devised an end-to-end product for telecom sector along with use case specific workflow which not only help in integrating analytics for better insight, but also lets one work through it without the need of a Data Scientist.
  - o Developed generic Vertica - R - User Defined Functions (R-UDFs) for various Statistical and Analytical Algorithms and generic workflows to address different use cases in telecom sector such as Churn and Next Best Offer.
  - o Developed the underlying data model and SQL transformation scripts for staging data, which can be plugged into existing client database for ease of application deployment.
  - o **Technologies:** R, Vertica, Spotfire

- **Automated pipeline for Data Quality using EDA – Scalable Analytics COE**
  - o Created an automated workflow and solution for data quality report using Exploratory Data Analysis
  - o A reusable data audit solution (product in itself) was created that could be used in potentially every project at the beginning to generate a DQ report on the data.
  - o **Technologies:** R

- **Flight Delay Prediction - Scalable Analytics COE**
  - o US Department of Transportation's historical flight data, Weather Data and Holiday Data was analyzed to figure out key drivers for delays in departure/arrival of flights.
  - o A regression model was developed to predict and generate triggers for possible delays in flights.
  - o The solution can be leveraged by the Transport industry to figure out and possibly avoid delays in flight. It can also be leveraged to assist passengers in planning travel by providing recommendation of routes, timings and possible flights.
  - o **Technologies:** R, Vertica, Hive, Spark

- **Performance Benchmarking of various R releases - Scalable Analytics COE**
  - o Publicly available data-sets from UCI and Stanford were used to sample and generate various fixed size data.
  - o The data was used for executing various analytical algorithms [Logistic and Linear Regression, Page Ranking, K-Means Clustering] over multiple iterations and the execution durations were recorded for benchmarking.
  - o **Technologies:** R releases used - Open Source R 3.1, Tibco Enterprise Runtime for R 3.1, HP Distributed R 1.0, Vertica R – UDF 7.1/3.0; TIBCO Spotfire

3. SYSTEMS ENGINEER
BFSI, Infosys Limited, Bangalore   **(*June 2013 – Jan 2015*)**

Business Intelligence, Data Warehousing and Big Data Engineer in the Banking Financial Services and Insurance vertical.

- **Account Performance Data Mart Development – Leading American Financial Service Provider**
  - o Development of a Big-Data Data-Mart for reporting the performance of Private Wealth Management accounts.
  - o Requirements gathering and analysis, and finalization of specifications
  - o Designing of the ETL Workflows, Collection designs, Capacity planning, DB designs for scalability and robustness
  - o Development and Unit Testing of individual modules [for ETL Workflows, aggregation SQL scripts, database controls, etc.]
  - o **Technologies:** Ab Initio Co>Op 3.1.7.2, Oracle 11g, Mongo DB 2.4.5, Oracle SQL Developer, NoSQL Mongo Client

- **Training – Data Engineering**
  - o Three months of rigorous training at Infosys Mysore in generic fields of Computer Sciences and specialized training on Business Intelligence/Data Warehousing, including Databases, ETL, Analytics and Reporting
  - o **Technologies:** JDK 1.6, MS SQL Server 2008, Oracle 10g, Microsoft BI Suite 2008, Eclipse IDE, IBM Cognos 10.1

SKILL SET

Technical: Data Warehousing, Data Modelling, Basic Understanding of Analytics and Machine Learning, SQL, NoSQL, Vertica, Hive, Oracle DB, MongoDB, Ab Initio, Spotfire, PowerBI, R, Apache Spark

Non-Technical: Consulting, communication and interpersonal skills, presentation, business requirement profiling, client handling, innovative thinking, leadership and team player

Certifications:
- Certificate in Network Management from Nettech – 2010
- Certificate in Web Based Enterprise Application Development using JEE from Globsyn – 2011

AWARDS & RECOGNITIONS

- Awarded multiple excellence awards for significant contribution to Projects at HP/HPE.
- Awarded for organizing FS CSI CARNIVAL at Infosys Limited
- Held the position of Vice-Secretary at the Institute of Electrical and Electronics Engineers (IEEE), HIT Student Branch during the session 2010
- Won multiple Coding competitions organized by IEEE 2009-2011.

INTERESTS & EXTRA CURRICULAR ACTIVITIES

Interested in music, travel, cooking, organizing events, social work and volunteering