Saptarshi Datta

Assistant Systems Engineer

Professional Member - [Association of Computing Machinery](http://member.acm.org/~saptarshidatta)

[Linked IN](https://www.linkedin.com/in/saptarshi-datta-2b5b70aa/) | [GitHub](https://github.com/saptarshidatta96) | [Website](https://saptarshidatta.in/home)

# Career Objective

# To work in a challenging and motivating career that provides an opportunity to work with the latest technologies.

# SKILLS

**Machine Learning:** Classification, Regression, Clustering, Decision Trees

**Deep Learning:** Forward Propagation, Backward propagation, Hyper parameter Tuning, Regularization & Optimization, CNN & RNN models. Experience with Keras API (Tensorflow Backend) snd PyTorch.

**Statistical Methods:** Predictive Analysis & Hypothesis Testing, Principal Component Analysis and dimensionality reduction techniques, Market Basket Analysis

**Programming Languages:** R, Python | **Big Data Language:** Hadoop

**Scripting Language:** Unix | **Database Language:** SQL

**Data Extraction Tool:** Informatica PowerCenter, IBM Infosphere DataStage

**Data Reporting Tool:** Tableau

# EXPERIENCE

## Tata Consultancy Services, Kolkata *- Assistant Systems Engineer*

November 2018 - PRESENT

**ROLE:** ETL Developer - Informatica PowerCenter

**Client: Confidential (A major Financial Services Company in BFSI Segment)**

● Assisted during the designing and development phase of complex ETL structures for transformation of data sources into data warehouses.

● Analysed database designs, system integration processes and Source to Target Mappings to make suggestions for further enhancements

● Implemented procedures for validation, reconciliation of metadata and error handling in ETL processes

● Resolved problems relating to ETL applications and data issues.

● Suggested improvements in processes for managing data integrity and quality for overall enterprise

● Designed and implemented processes for manual loading of catch up loads including scheduling and unscheduling of jobs taking care of dependencies between various jobs.

# PROJECTS

## Creation of an India Credit Risk Default Model Using Logistic Regression

The project involved developing a credit risk default model using a given data which had to be checked for outliers, missing values, multicollinearity etc. Univariate and Bivariate Analysis had to be conducted and the model had to be built using Logistic Regression on most important variables. Model Performance Measures were undertaken that included predicting the accuracy of the model on certain datasets.

**Tools**

Logistic Regression, Univariate & Bivariate Analysis, Outlier Treatment, Model Performance Measures

## Visualizing Car Insurance Claims using Tableau

This project explored the art of problem-solving with the aid of visual analytics. Tableau’s data visualization tools were used to create interactive dashboards to provide high-level insights to an Insurance company to drive the company's car insurance schemes.

**Tools**

Data Visualization, Tableau, Business Intelligence

# CERTIFICATIONS

* Python for Data Science and Machine Learning Bootcamp, Udemy
* Deep Learning Specialization, Coursera

**EDUCATION**

**Great Lakes Institute of Management**

Post Graduate Program - Business Analytics & Business Intelligence

(May 2019 – May 2020)

**Maulana Abul Kalam Azad University of Technology, West Bengal**

Bachelor’s Degree in Electrical Engineering, DGPA: 8.46/10

(August 2014 – June 2018)