

# Sushmita Bhattacharya

## Areas of Interests

Machine learning, Data Sciences.

## Professional Experience

- Working as a **Software Engineer** in **Microsoft India Development Center**.
- Data Scientist in **Honeywell Technology Solutions Labs, India**
- Worked as a developer in **Cognizant Technology Solutions, India**

## Academic Background

Standard	Institutions
M.Tech	Indian Institute of Technology Bombay
B.Tech	Bengal Engineering and Science University Shibpur, Howrah

## Key Projects

- **Anomaly detection and Root cause analysis:**
  - **Objective:** Anomaly detection and root cause analysis for time-series in search ads scenario.
  - **Responsibility:** To automating hugely tedious manual efforts of finding root cause of anomaly in search ads framework. Significant deviation from normal trend of the series (slice in a multi-dimension data warehouse) flagged as anomaly. Smart pruning techniques and domain knowledge used to decrease search space in large OLAP database. Root cause detector focused on finding out major contributing dimension members and metrics which were correlated and indicated significant shift while the event occurred based to their magnitude of importance. Built Excel plug-in to surface investigation summary to the end user.
  - **Methods used:** Neural Network, Autoregressive integrated moving average, Tree augmented Bayesian Network, Sub-sequence clustering.
  - **Technology:** C#, MS Excel.
- **Mold Press Analytics:**
  - **Objective:** Molding process failure prediction for manufacturing breaks.
  - **Responsibility:** Cleansed and preprocessed data from readings of various sensors attached to the surface of container. Identified important features. Compared and evaluated different machine learning models (Discriminant Analysis, Logistic regression, SVM, Neural Network) built on the data. The classifier's accuracy perfected using sub-sampling to eliminate data skew.

- **Technology:** R, Tableau.
- **Analytics on Purchase Order:**
  - **Objective:** Purchase order arrival prediction for better productivity of the organization.
  - **Responsibility:** Preprocessed data, detected important features and evaluated different models (Decision Tree, Naive Bayes etc) built on the dataset. Explored how the various of machine learning algorithms can be scaled in a distributed environment (using spark MLlib).
  - **Technology:** R, Weka, Apache Spark, excel.
- **Big Data Analytics in a Distributed Systems Environment** (Master's thesis)  
(Guide: Prof. N. L. Sarda)
  - **Objective:** Outlier Detection in Big data
  - **Approach:** Stored large amount of incoming data in a distributed store(HDFS). Built offline model on Hadoop system and predicted outliers in stream processing system (Apache Storm, HBase). Explored various tools and techniques for storing and processing big data and data mining methods for outliers detection.
  - **Technology:** Hadoop HDFS, Map-reduce, Apache Storm, HBase, Redis Queue, Gnuplot etc.
- **Streaming Data Processing and Management** (Jan 2014 - May 2014)  
(Guide: Prof. N. L. Sarda)
  - Studied streaming data and its difference from traditional relational data and processing.
  - Surveyed stream query language and special purpose storage and indexing for streams.
  - Reviewed STREAM - a Stanford implementation for stream data management system.

## Academic Projects

- **Geometry Generalization for Map Simplification** (Guide: Prof. N. L. Sarda in Spatial Database)
  - Performed simplification of the linear geometries without affecting the topology of geometries using modified Ramer-Douglas-Peucker algorithm in **Java**.
  - Achieved runtime of 300 ms to simplify a set of linestrings with 900 data points.
- **Part of Speech Tagging** (Guide: Prof. Pushpak Bhattacharya in Natural Language Processing)
  - Developed a part of speech tagging system for English sentences in **Java**, with an average precision of 93%.
  - Implemented **Viterbi algorithm** for **Hidden Markov Model**.
- **Implementation of Table Partitioning in PostgreSQL** (Guide: Prof. S. Sudarshan in Implementation Techniques for Relational Database Systems)
  - Modified source code of PostgreSQL to gain table (range) partitioning functionality
  - Changed code for insert, delete and update of tuples to take place in proper partition.
  - Added code for creating index(s) in the partitioned tables if one is present in main table.
- **Color and Size Based Fruit Sorter using FireBird V** (Autumn, 2013)  
(Guide: Prof. Kavi Arya and Prof. Krithi Ramamritham in Embedded and Real Time Systems)
  - Built modular hardware and software components for feature based, real time fruit sorter.
  - Coded various sensors and actuators in Firebird V. Written code for serial communication between Firebird V and PC.
  - Designed and performed experiments with various test-sets and got **90% accuracy**.

## Skill Set

- *Technology and Tools:* Machine learning tools and techniques, (used R, scala), Big Data tools (Hadoop, Map-reduce, Hive, HBase, Apache Storm, Spark)
- *Programming Languages:* Core Java, Go, C, C++, C# PL/SQL, SQL Server
- *Scripting Languages:* Python, Bash
- *Operating Systems:* Linux, Windows
- *Other Tools:* Weka, L<sup>A</sup>T<sub>E</sub>X, Eclipse, Gnuplot

## Positions of Responsibility

- **Teaching Assistantship:** Involves mentoring students, conducting quizzes, grading assignments and exams.
  - Embedded Systems Lab (Spring, 2014)
  - Database and Information Systems Lab (Autumn, 2014)

## Achievement and Extracurricular activity

- **Academic Achievement:** Secured All India Rank of 57 amongst 2,24,160 candidates appeared in Graduate Aptitude Test for Engineering, 2013 (in Computer Science stream) .
- **Hobby:** Painting