Chandan Saha

chandan.saha@rutgers.edu | 857-800-4211|

EDUCATION

RUTGERS UNIVERSITY, NEW BRUNSWICK , NJ
MASTER'S IN DATA SCIENCE AND ANALYTICS
May 2017

VIT UNIVERSITY
BACHELOR'S IN COMPUTER SCIENCE
June 2011 | Vellore, IND

SKILLS

PROGRAMMING
Python • C++• Scala

DATA SCIENCE
Machine Learning • Deep learning • Regression Analysis

DATABASES
SQL • Hive • NoSQL • Elasticsearch

BIG DATA
Hadoop • Map/Reduce • Apache Spark • Kafka

STATISTICAL TOOLS
R • Matlab• SAS

MISCELLANEOUS
AngularJS • Node.js • D3.js
Git

COURSEWORK

GRADUATE
Machine Learning
Pattern Recognition
Regression Analysis
Linear Algebra
Fundamental of Analytics
Big Data and Cloud Computing
Applied Statistics

UNDERGRADUATE
Algorithm and Data structure
Artificial Intelligence
Advance Database Management

LINKS

Github:// **Chandan**
LinkedIn:// **Chandan**
Website:// **Chandan**

Work Authorization
Relocation : Yes
F1-OPT : 36 Months

PROFESSIONAL EXPERIENCE (4+ YEARS)

CAIT | RESEARCH ASSISTANT - DATA SCIENTIST INTERN
October 2016 - Present | Rutgers University , Piscataway , NJ
- Building quantitative / Predictive modeling (**feature selection, survival probability and validation metrics**)for Long-Term Bridge Performance project to do Survival analysis on national bridges in United States.
- Technology Used : **R , Python** , SQL Server , Java Script

RUTGERS DISCOVERY INFORMATICS INSTITUTE | DATA ENGINEER INTERN
June 2016 - August 2016 | Rutgers University, Piscataway , NJ
- Benchmarking of in-memory processing for **Clustering and optimization of large Graph Database** in distributed machines.
- Technology Used : **Ingestion** : Kafka, **Storage** : Hadoop(hdfs) Batch Processing: **Apache Spark** - Geo-spatial search ,**Database** : Cassandra

INTEL CORPORATION | SOFTWARE DEVELOPER
June 2011 – July 2015 (4 + years) | Enterprise Development Team
- Led a team of 3 and developed for China material handbook project for Intel china and vietnam users with **saving of $ 500k/year** .
- Designed complex business changes events for forecasting inventory and logistic overhead for Intel customers to **reduce overhead cost by 5%.**
- Building and mainitaining **Data pipeline , ETL processes** for for various ERP applications across Intel IT.
- Technology used: **SQL,Java,JavaScript(AngularJS)** , SAP , SAP HANA , HDFS , BI Reporting , middleware technologies.

INTEL CORPORATION | SOFTWARE DEVELOPER INTERN
September 2010 – June 2011 | Sales and Marketing IT
- Developed and upgraded features of **Customer Relationship Management** application , a customer facing Web Application for Intel Warranted Products Issue resolutions. Technology used: **SAP , Java** , HTML, Adobe Flex.

PROJECTS

- Developed **Real-time Traffic Sign Classifier** for self-driving cars using **Convolutional Neural Network** using Caffe and Python . Improved accuracy by $\approx$ 3% from state-of-art model.
- Built a **scalable data pipeline** to recommend parking spots in real time based on the parking sensor data of San Francisco city - EasyPark.
- Developed **Dynamic topic models** for monitoring business competition of Yelp restaurants using NonNegative Matrix Factorization, STC and LDA.
- Constructed logistic regression pipeline to predict Click Through Rate on **Kaggle competition** data based in Apache Spark .

ACHIEVEMENTS

- Winner of the **Social Initiative Contest** for cancer patients project at Intel
- Winner of Code-A-Thon, a **Hackathon** organized by Intel Labs.
- Received **Asia Finance Divisional Recognition Award** to protect Intel's Duty Free Status in China.
- Received **GATE scholarship** and taught computer science to under graduates and at Vellore Institute of Technology.
- Berkeley **Big Data XSeries - Apache Spark** certification.