

Tanay Chowdhury

Mobile: (818) 975 -0087

Email: tanaychowdhury@yahoo.com

 profile : <http://linkedin.com/in/tanayz/>

Mailing Address

928 W 36th St, Apt BR,

Chicago, IL 60609,USA

Website: <http://tanayz.com>

EDUCATION:

Master of Science in Computer Science

University of Illinois at Chicago, GPA 3.28

Completion: May, 2014

Bachelor of Technology in Information Technology

Kalyani Government Engineering College, GPA 7.9/10

Completion: July, 2005

Related Coursework:

1. Data Mining
2. Machine Learning
3. Natural Language Processing
7. Machine learning course from coursera

4. Mathematical Statistics
5. Advanced Data Mining
6. Data Semantics

CERTIFICATIONS

- **Brainbench** certified in Business communication
- **Cognizant** certified Data warehousing Professional.
- **Oracle** certified Professional in Oracle 9i Sql.
- **IBM** certified in Oracle as an RDBMS.
- **Informatica** certified in developer and architect module.
- **Big Data University** certified in **Hadoop** fundamental.

EMPLOYMENT

Data Scientist, Microsoft Mobile Oy

April 2014 – Present

- Use sql to query, roll up and then pull data from Amazon or Hadoop.
- Analyze and build data mining and Statistical model using R, Python.
- Fit Arima (p, d, q) time series model to predict future daily and weekly business results.
- Reporting and visualizing data to make drill down dashboards using tableau.
- Enhance and perform A/B testing of Java map-reduce based recommendation system.

Data Analyst, University of Illinois at Chicago

June 2013 – April 2014

- Analyzed functional requirements and created business object infoview report on the basis of that.
- Analyzed business requirement to create visual statistical report using Tableau.
- Used SAS and R to do **linear and logistic regression** and clustering on test data and build distance matrix.
- Develop and enhance web transport portal using PHP and MySQL.

BI Consultant & Data Analyst, Cognizant Technology Solutions

October 2005 – January 2013

- Utilized analytical applications like **SAS,R** to identify trends and relationships between different pieces of data
- Mined large health care databases to research internal requests using Java and R.
- Designed and implemented **statistical / predictive models** to predict demand and enrolment in insurance.
- Designed data model for enterprise data warehouse capturing functional requirements from users.
- Developed and implemented end-to-end data warehousing solution using **Informatica**.
- Developed Unix shell script for source data validation and load.
- Created business report using Business object.
- Created **Proof of Concept (POC) on using Map-Reduce framework** on healthcare and financial data.
- Successfully implemented DW solutions at blue-chip clients like Pfizer, JP Morgan Chase and Credit-Suisse.

ACADEMIC PROJECTS

Master's Thesis

Enhance the accuracy of classifiers on sparse and high dimensional fMRI data with the help of dimension reduction, smoothing and algorithmic feature selection, bagging and boosting approach.

Multi-label Crime Forecasting

Designed intelligent spatial and temporal feature and applied **tree based probabilistic meta chain classifier** on crime records to predict probability of criminal activity and its intensity on a certain place with very high accuracy(92.2%).

Sentiment Mining

Developed a twitter sentiment analyzing classifier system to classify positive and negative tweets made during the 2012 presidential election using Matlab 2012b for data cleanup and Weka 3.7.9, and Rapid Miner 5 for doing **feature pruning** and creating **TF-IDF word vector** and obtain **highest** accuracy (66.42%) and F-score in the class of Data Mining.

Next Stay location predictor

Used GeoLife GPS Trajectory data collected by Microsoft Research Asia to build a probabilistic suffix tree model to predict next stay location in the trajectory. The best result which was obtained optimizing the support value and eps was **better** compared to Markov Chain (k-th order markov) model.

Automatic essay grader

Developed an **automatic essay grader** judging word order, subject-verb agreement, verb tense / missing verb / extra verb, coherence, topic following and length of the essay using java in eclipse with Stanford parser, tagger and wordnet library with **3rd** highest accuracy in the class of natural language processing.

TECHNICAL SKILLS

- **Operating Systems:** UNIX, Linux, WINDOWS;
- **Tools and Utilities:** MS-Office,
- **Programming Language:** Core Java, PHP, Java Script, SQL, PL/SQL, Python;
- **Statistical Packages:** MATLAB, R, SAS, Tableau, Weka 3.7.9, Rapid Miner.
- **Database(s) :** Oracle 8i/9i /10g,IBM DB2,MySQL,Microsoft SQL Server
- **Datawarehousing Tools :** Informatica Powercenter 8.6, Datastage 7.51,SAP Business Object 4
- **Big data tools :** Hive, Map-reduce, NoSQL (MongoDB)