Mikit Kanakia

(412) 609 – 0624
mikitkanakia@gmail.com
https://mikitkanakia.github.io/
https://www.linkedin.com/in/mikitkanakia

EDUCATION

Carnegie Mellon University, Pittsburgh, PA — May 2017
Master of Information Systems Management
<u>Relevant Courses:</u> Data Mining, Big Data and Large-Scale Computing, Data Warehousing, Text Analytics, Programming R Analytics, NoSQL Databases, Python for Developers, Distributed Systems

Veermata Jijabai Technological Institute, India — Jun 2014
Master of Technology in Computer Engineering – Software Engineering

University of Mumbai, India — Jun 2009
Bachelor of Engineering in Computer Engineering

EXPERIENCE

Carnegie Mellon University, Pittsburgh, PA — Sept 2017 – Present
Research Intern, Python, R
- Analyzed data from wearable devices and segmented the medical symptoms using cluster analysis

Amdocs India Pvt. Ltd., India — Aug 2014 – Apr 2016
System Analyst, SQL, PL/SQL, Python
- Employed **predictive analytics** and learned the usage patterns of the consumer for cross selling opportunities
- Automated bundling for 5 Million products with **zero defects** using PL/SQL for Vodafone India
- Designed ETL process and data migration plan for 68.5 million consumers of XL Axiata
- Performed exploratory data analysis on usage and billing data for 200+ production critical issues

University of Mumbai, India — Feb 2010 – Sep 2012
Lecturer
- Delivered several courses including Database Management Systems and Software Engineering

SKILLS

Programming: Python (numpy, scipy, nltk), R, SQL, PL/SQL, Java, C, C++, Unix Shell Scripting
Databases: Oracle 11g, MS SQL Server 2014, MongoDB, MySQL
Analytics Tools: Tableau, Octave, MS SQL Server Management Studio 2014 (SSIS, SSAS), Weka
Big Data: Spark, Hadoop, Hive

ACADEMIC PROJECTS — May 2016 – May 2017

Location Prediction of Businesses in Yelp Dataset
- Developed a classification model and determined the location of the 1000+ businesses using **MLlib** in Spark
- Performed feature engineering on **~25 GB of semi-structured data** and optimized the model performance

Click-through Rate Prediction Pipeline
- Developed Logistic Regression Model for Click-through Rate prediction for online ads
- Reduced feature dimensions using feature hashing on categorical data using One-Hot-Encoding

Human Activity Prediction from Sensor Input
- Predicted position of body activity by analyzing time series data from various accelerometers using R
- Applied various machine learning algorithms like Decision Trees, k-NN for classification of the human activity

Analyzing U.S. Patent Applications using Python
- Created an online dashboard for keyword extraction using **Natural Language Processing (NLP)**
- Performed web scraping, noise removal, scrubbing, stemming, normalization and word tagging of keywords

Degree Distribution Computation of Friendster Dataset
- Calculated in-degree and out-degree of 10 Billion nodes using Hadoop framework and Spark GraphX

Data Warehouse for Drought Data Processing
- Migrated the data with the size of ~475K rows to data warehouse using **ETL** technologies
- Prepared star schema and performed dimension modelling on the data and provided business insights

ACHIEVEMENTS AND CERTIFICATIONS

- Finalist in Internet of Things(IoT) Idea-thon on - **Exploring Data for Wearable Devices** — Mar 2017
- **Certificate of Excellence** for achieving significant milestones, Vodafone India Project — Feb 2015
- **Oracle Database: SQL Certified Expert**, (Oracle) — Jan 2011