

UTKARSH KUMAR SINGH

3315 E Longview Avenue, Bloomington, IN 47408

utksingh@iu.edu | Tel: (812) 955-1508 | www.linkedin.com/in/u-singh | www.github.com/Utkarsh101

EDUCATION

Indiana University (IU), Bloomington, IN, United States
Masters of Science in Data Science

GPA: 3.7/4.0

Aug 2016 - Present

Relevant Coursework: Data Mining, Advanced Database Concepts(SQL), Introduction to Statistics, Exploratory Data Analysis in R, Data Analysis & Modeling in SAS, Social Media Mining(Python)

National Institute of Technology (NIT), Durgapur, India
Bachelor of Technology, Information Technology

GPA: 8.14/10

Aug 2012 - May 2016

PROFESSIONAL EXPERIENCE

Westfield Gas & Electric, MA
Data Analyst Intern

Summer 2017

- Worked on **Smart Meter Data Analysis** project whose aim was to analyze smart meter data and get important insights on electricity conservation measures and initiatives for utility as well as customers.
- Installed the latest stable version Apache Hadoop-2.8.0 on the Oracle VM Virtual box and its other dependency softwares in the production environment.
- Completed batch ingestion of smart meter data into HDFS using a flume agent on Apache Flume 1.7.0.

National Informatics Centre (NIC), Delhi
Summer Trainee

Summer 2015

- Worked on the Project **Manpower Management & Tracking System** whose aim is to track the people working within an organization.
- Developed a module to improve the efficiency of the organization by optimally utilizing the resources (manpower).

Informatica Business Solutions Pvt. Ltd., Bangalore
Marketplace Intern

Summer 2014

- Worked on **PowerCenter & PowerCenter Express**, the flagship tools of Informatica (ETL tool used for extracting data from various sources, transforming & loading data into multiple targets) & got acquainted with them.
- Developed Python scripts(for both windows & Linux) to enhance the utilities of PowerCenter:
 - Create relational connections (Oracle, DB2, ODBC, Sybase & Teradata) in bulk to establish connectivity between databases & Informatica repository.
 - Check validity of the relational connections after creating them.
- Informatica Marketplace links where the above-developed scripts published:
 - https://community.informatica.com/solutions/python_create_bulk_connection_for_powercenter
 - https://community.informatica.com/solutions/python_create_test_powercenter_connections

TECHNICAL SKILLS

Programming Languages: Python(Scikit, numpy, pandas), R, SQL, SAS, C, C++
Databases: SQL Server : T-SQL, SSMS, SSIS, SSRS, MYSQL, PostgreSQL, Oracle
Big Data Technology: Hadoop, Spark, Flume
ETL Tools: Informatica PowerCenter, Informatica PowerCenter Express
Data Visualization: Tableau, ggplot2 and ggvis (R), matplotlib (python)
Tools and Miscellaneous: R Studio, R Markdown, Latex, HTML, Microsoft Visual Studio

RELEVANT PROJECTS & COURSES

- Currently working on the project "**Data Analysis of Cryptocurrencies**" to analyze Bitcoin & 1200+ cryptocurrencies trends using PostgreSQL in order to get high-level insights into the cryptocurrency market and predict the evolving market trends.
- Created a predictive model to predict **Stock Market movements of 'Nike'** using supervised machine learning Principles to the tweets extracted from Twitter and developed a sentiment analyzer that can judge the type of sentiment present in the tweet. Analyzed the correlation between stock market movements and sentiments in tweets.
- Created a predictive model to accurately predict first booking destination for new Airbnb customers as part of the **Kaggle competition "Airbnb New User Bookings"** using advanced ML algorithms such as xgboost & secured a rank in top 20% of the competition.
- **Analytics & Visualization Challenge** - Performed data analysis & build visualizations using Tableau that uncovered significant trends in data and identified factors that could influence the sales and increase revenue of grocery stores across various demographics.
- Secured a rank in top 20% of a "**Kaggle competition: House Prices Prediction**" while building a model to predict the prices of residential homes in Ames, Iowa, using feature engineering and advanced regression techniques.
- **Coursera Certifications** - Introduction to Big Data Analytics - UCSD. Data Science Specialization - Johns Hopkins University.
- **SQL Server Administration Certification** - Implemented real world Production environment techniques such as, **backup and restore procedures**, monitor health of SQL Server, review storage, data and log file necessities, configuring security & setting up maintenance plans, Performance tuning using **SQL Profiler, Database Tuning Advisor(DTA) & DBCC scripts**, create & investigate issues using **Database Management views & extended events**, configure & monitor **High Availability Solutions** such as **log shipping, database mirroring, replication & Always on availability** group technology, deploy **SSRS** for reports, deploy **SSIS** packages to migrate data

LEADERSHIP

- Captained the Table tennis team of National Institute of Technology (NIT) Durgapur in Inter-NIT tournament.