YASHASWINI DHATRIKA

LinkedIn: https://www.linkedin.com/in/yashudhatrika/ |**GitHub**: https://github.com/YashuDhatrika

Mobile:(812)-325-4634 | **Email:**ydhatrik@gmail.com

Education

Master's Degree in Data Science   **(Aug 2018-May 2020)**
Indiana University Bloomington.   *GPA 3.81/4.00*
Coursework: Machine Learning (ML), Elements of Artificial Intelligence (AI), Probability & Statistics, Time Series and Signal Processing
Machine Learning in Computational linguistics (NLP), High-Dimensional Data, Deep Learning Systems and High-Performance Big Data Systems.

Bachelor's Degree in Textile Technology   **(Aug 2011-May 2015)**
Indian Institute of Technology (I.I.T.), Delhi, India.
Relevant Coursework: Cluster Analysis in Data Mining, Mathematics I (Differential Calculus), Mathematics II (Vector Calculus and Complex Analysis) and Managerial Economics.

Professional Experience:

Data Scientist | Hatcher+   **(July 2020-Present)**
- **Developed** an algorithm which generates the vectors for customized textual data using sentence embedding and word /character level embedding models (Elmo, BERT, USE, Inferscent, etc.)

Data Analyst | Kelley School of Business, USA   **(June 2019 – May 2020)**
- **Developed** a couple of workflows to analyze the performance of the various campaign in **Alteryx** that inputs the data of various sources from Google Analytics, Vendor Database, and Salesforce and then pushes the cleansed and transformed data to the MS SQL Database.
- Performed channel attribution **Markov model** to identify the effective channels and paths contributing to conversions.

Revenue & Planning Analyst | IndiGo Airlines, India   **(May 2017 – July 2018)**
- Developed a **forecasting engine** using LSTM to forecast the demand of the passengers for ~1000 flights which operated ~70 destinations. **The optimized engine** showed the Revenue (Passenger revenue per available seat kilometer) a **7.7% increase** for 6 months compared to the prior year for the same period.

Business Analyst | BRIDGEi2i Analytics Solutions Private Limited, India   **(Jun 2015 – May 2017)**
Awards: Awarded the Best Team award **twice** for a period of 2 years for providing the quality & impactful insights for the client.
- **Developed** a propensity model (Logistic Regression) to prioritize the customers and identify effective means of communication for cross-selling across multiple products for a financial company. The impact of the implementation showed improvement in the cross-sell conversion by **5%** compared to prior months
- **Analyzed** online purchase intentions and **developed** a propensity model (Logistic Regression, Spark) using clickstream data for a technology firm to identify the best prospects for targeted marketing efforts. The impact of the implementation showed improvement in the conversion by **13%** compared to prior months.

Research Experience:

Research Assistant |Indiana University Bloomington   **(Jan 2019-Aug 2019)**
Prediction of Gestational Diabetes in Pregnant Women:
Tools /Concepts: Python, Expectation–maximization algorithm and multiclass logistic regression.
- **Developed** a baseline model (SVM, Logistic Regression, Random Forest) and a mixture of an expert model (probabilistic model) to predict gestational diabetes in pregnant women.
- Improved accuracy from **72% to 93%** by adopting a mixture of an expert model over other baseline models.

Projects:

Customer Profiling & Lead Prioritization Modeling:
Tools/Concepts: R, SQL, Logistic Regression.
- Built scorecard models using logistic regression to prioritize leads and efficiently allocate resources for marketing campaigns.
- Helped the client in reducing the acquisition cost of new customers up to **11%** of the prior year using purchased data.

Classification algorithm to determine in-risk flights:
Tools/Concepts: R, Hive, Random forest, Adaboost and SVM
- Identified risk flights using random forest based on current bookings, remaining seats, booking velocity, etc. The algorithm helped the flight analyst to put the relevant fare points in the reservation system and helped in boosting the PRASK (Passenger revenue per available seat kilometer) by **3.7%.**

Implementation of Text to SQL for NLIDB systems:
Tools/Concepts: Python, NLP, Probabilistic Model (Hidden Markov Models, Maximum Entropy Markov models and Conditional Random fields) & Deep Learning Models (Encoder-Decoder Architecture and the Attention Based architecture)
- **Developed** a system based on a probabilistic and deep learning approach that can yield SQL queries automatically by simply interpreting the natural language queries.
- Achieved **accuracy of ~80%** by deploying the deep learning model.

Technical Skills

- **Languages:** Python (Pandas, NumPy, Scikit-learn, TensorFlow, Keras, NLTK, GloVe, spaCy, bokeh, seaborn), MATLAB, R, SAS, SQL, Java & JavaScript
- **Databases and Big data -** MySQL, MS SQL Server, PostgreSQL, Databricks, Hadoop, Spark, and Hive.
- **Other tools:** Tableau, Microsoft Azure, Alteryx, Power BI, Adobe Analytics, Google Analytics, and Advanced Excel
- **Competencies**: Forecasting, Probabilistic Models, Statistical modeling, Regression, Classification models (Decision Trees, Bagging and Boosting, SVM, Logistic Regression), Clustering (Hierarchical, K-means), A/B Testing, Natural Language Processing (NLP) and Deep Learning frameworks (LSTM, CNN, Seq2Seq)