# DIMPLE BAPNA

bapna.d@northeastern.edu | GitHub | LinkedIn | Portfolio | Seattle, WA | (425) 233-7801

## EDUCATION

**Northeastern University**, Master of Science, Data Analytics Engineering **GPA: 3.8** Aug 2019 - Jan 2021

*Courses*: Data Mining, Neural Nets and Deep Learning, Data Warehousing and Business Intelligence, Database Management & Design, Data Visualization, Probability & Statistics, Deterministic Operations, Statistical Methods in Engineering

**University of Mumbai,** Bachelor of Engineering, Computer Engineering **GPA: 3.8** Aug 2008 - Jun 2012

## EXPERIENCE

**Data Science Consultant** **Costco Wholesale, Seattle** Jun 2020 – Aug 2020

- Implemented LSTM to forecast weekly demand of goods sold by Costco.com, resulting in reduction of revenue loss by 20%
- Developed comprehensive visualizations depicting forecasted demand at depot-level using Tableau dashboard
- Leveraged data exploration techniques to identify limitations and recommended improvements in data quality

*Tech stack: Python, matplotlib, seaborn, sklearn, neural network(LSTM), tableau*

**Data Science Consultant** **Tignis, Seattle** Feb 2020 – Jun 2020

- Implemented anomaly detection for real-time sensor data with neural networks (Feed forward) by utilizing historical data
- Performed statistical modeling on sensor data for data manipulation and missing value treatment
- Reduced the overall time taken for fault investigation by 60%, resulting in direct cost savings of more than $300k for a Fortune 500 client

*Tech stack: Python, matplotlib, sklearn, neural network, statistics*

**Senior Software Developer** **Duck Creek technologies, India** Aug 2016 - Jun 2018

- Led a team to build a rule-based billing system resulting in a 35% faster billing process for a US insurance giant
- Developed an automated payment records system compatible with 50+ banks, saving 160 man-hours per month
- Improved payment experience by driving a data-informed policy change, reducing customer attrition rate by 20%

*Tech stack: C#, SQL, Duckcreek tools*

**Software Developer** **Accenture Services Pvt Ltd, India** Feb 2013 - Jul 2016

- Automated medical billing application for incident analysis, increasing productivity of the healthcare system by 70%
- Developed a single input application which eliminated manual interventions reducing engineering costs by over 60%
- Delivered a tool to eliminate duplicates and highlight pending invoices saving $2M+ in interest penalties for the client

*Tech stack: C#, SQL*

## DATA SCIENCE PROJECTS

**Research Thesis: Image classification using Convolutional Neural Networks (CNN, TensorFlow, Python)** May 2020 - Oct 2020

- Designed a Convolutional Neural Network for image classification using CIFAR-100 dataset
- Improved the accuracy of the model using various optimization techniques like regularization, image augmentation, etc.
- Achieved an accuracy of 48.39% using only 3 fully connected layers and 4 convolution layers

**ETL pipeline for immigration data (AWS Redshift, Apache Airflow, Spark, ETL, Python)** Jun 2020 - Aug 2020

- Built a data lake on S3 to store information from various data sources and run analytics on US immigration data
- Designed end to end ETL pipeline extracting data from S3 buckets into staging tables and modelled the data into star schema
- Used Apache Airflow to automate ETL pipeline, utilizing Python and AWS Redshift

**Predicting customer attrition using historical data (Python, Sklearn, TensorFlow, Keras, Neural Network)** Mar 2020 - May 2020

- Trained a neural network to predict whether the borrower will pay back the loan, using historical data on loans
- Analyzed the data and performed feature extraction and feature engineering to fit it to the neural network
- Achieved an accuracy of 88% by adding dense layers, and implemented early stopping, to avoid over fitting

**Automated grocery list generator (R Shiny, Python, ML- Apriori)** Feb 2020 - Apr 2020

- Developed and deployed R Shiny web app to break down recipe and generate grocery list leveraging a dataset of 1M Recipes
- Performed data transformation and data wrangling to extract ingredients and nutrition values
- Applied Apriori algorithm to generate 27000 association rules and provide recommendation based on items in Grocery List

**Machine learning based SMS message filter (Python, Sklearn, NLP, Multinomial Naïve Bayes algorithm)** Nov 2019 - Dec 2019

- Processed dataset of raw SMS messages by data wrangling, cleaning and performing feature selection
- Applied Natural Language Processing and Multinomial Naïve Bayes machine learning algorithms, to create a spam detection filter to filter out unwanted and abusive messages for a user
- Achieved a precision of 99% on the test data

**Statistical analysis on healthcare dataset from Kaggle (R, Statistics)** Nov 2019 - Dec 2019

- Synthesized Kaggle healthcare dataset to extract insights and attributes leading to heart strokes in humans by performing data cleaning and exploratory analysis
- Conducted various statistical tests on datasets including null hypothesis, z-test, t-test, Chi-square goodness-of-fit test.

## SKILLS

**Machine Learning Algorithms:** Linear Regression, Logistic Regression, Decision Trees and Random Forests, KNN classifier, K-Mean Clustering, SVM, Apriori, Neural Nets and Deep Learning

**Languages:** Python (Pandas, NumPy, SciKit-learn, TensorFlow, Keras, PyTorch, matplotlib, seaborn, ggplot), R, Hive, SQL, Java

**Applications & Tools:** SQL Server, PL/SQL, Postgres, MongoDB, Jupyter Notebook, Tableau, R Studio, R Shiny, Git, Hadoop, IntelliJ, JIRA, Bitbucket, Confluence, Lingo solver