

# BAVITHRA ARUMUGAM

Email: [ba365@njit.edu](mailto:ba365@njit.edu) Mobile: 201 467 9390

## EDUCATION

### NEW JERSEY INSTITUTE OF TECHNOLOGY

Master's in Data Science

Aug 2019 - May 2021

GPA - 3.7

### ANNA UNIVERSITY, INDIA

Bachelor's in Information Technology

Aug 2013 - May 2017

GPA - 3.7

## PROFESSIONAL EXPERIENCE

### ICREDITWORKS, NEW JERSEY - DATA SCIENCE ASSOCIATE

Jun 2021 - Present

#### DATA ENGINEERING AND ANALYTICS

- Proposed, designed, implemented OLAP solution in Redshift and automated data ingestion using Glue and Lambda
- Remodeled analytics data pipeline and automated entire pipeline using Airflow.
- Migrated existing reports to the new ecosystem.
- Developed interactive dashboard for Application Analytics focusing business KPIs using Power BI.

### CBS INTERACTIVE, NEW YORK - MACHINE LEARNING INTERN

Jun 2020 - Sep 2020

#### CLOUD COST PREDICTION

- Developed and tested ML services to automate Cloud Cost Prediction and Optimization for AWS and GCP cloud usage across 40+ organizations, 100+services, 14 BUs by performing time series analysis.
- Executed the end-to-end project implementation, from requirement analysis to testing and deployment.
- Slashed human time and cost spent on prediction by 95%, ensuring the prediction accuracy and quality.

### NEW JERSEY INSTITUTE OF TECHNOLOGY, NEW JERSEY - RESEARCH ASSISTANT

Sep 2019 - Feb 2020

#### SOCIAL MEDIA ANALYSIS - CROSS PLATFORM COORDINATION

- Developed an ML model using NLP in Keras that predict whether a Twitter account is a troll within a set of 170K control accounts.
- Demonstrated usage of this model to find active accounts on Twitter still likely acting on behalf of Russia.
- Model is extendable to identify Russian troll accounts across platforms like Facebook, Reddit, etc.,

### RADIAL OMNICHANNEL TECHNOLOGIES, INDIA - ASSOCIATE SOFTWARE ANALYST

Jun 2017 - Jun 2019

#### DATA ENGINEERING AND INTEGRATION

- Developed integration solution between the Radial's E-commerce products to multiple clients using WebMethods by analyzing data of large volumes.
- Developed application design, including data modeling, data transformation and performance efficient dashboard development across global projects.
- Upgraded WebMethods Integration server from 7.1 to 9.6 version and performed end-to-end testing of modules.

### HOSTNET MEDIA, INDIA - APPLICATION DEVELOPER INTERN

Feb 2017 - Apr 2017

- Designed a mapping application that allows users to view a real time stream of Twitter hash tags using Node.js and D3.js

## PROJECTS

- ADVERTISING CHURN PREDICTION MODEL** - Developed Gradient Boost Classifier to predict advertising churn with 97% ROC-AUC.
- QUANTUM MACHINE LEARNING** - Developing auto hyperparameter tuning and optimization for Machine Learning models using Quantum bits using the properties of Quantum Superposition and Quantum Entanglement.
- AWS DEEP RACER ON REINFORCEMENT LEARNING** - Trained an RL model on AWS DeepRacer in a simulated environment, optimized the reward function to detect and avoid obstacles placed on the track and to finish in the best lap time.
- COMPARING INFORMATION SHARING BEHAVIOR ACROSS SOCIAL PLATFORMS** - Developed an ML model to classify between troll and non-troll accounts with a precision of 79% and an AUC of 96.5%.
- TIME SERIES ANALYSIS OF SALES AND INVENTORY** - Developed an ML model to predict the product sales at any given period to facilitate inventory management using Prophet.
- CANCER SURVIVAL TIME PREDICTION OF BRAIN TUMOR MRI IMAGES (Patentable Subject)** - Developed an ML model to predict the cancer survival time using BraTS2018 image dataset with ResNet with 67% test accuracy (65% is the highest achieved).
- SLEEP SCORING ANALYSIS USING BRAIN SIGNALS** - Developed an ML model using ResNet for 5 stage sleep scoring on PhysioNet Sleep-EDF (EEG, EMG, EOG signals) dataset and achieved 94% test accuracy.
- PORTFOLIO MANAGER USING R** - Developed a portfolio manager using R to select the best investment ETFs based on risk, returns and duration parameters.
- DIABETES PREDICTION USING K NEAREST NEIGHBOR** - Developed a classifier model to predict diabetes in patients using PIMA diabetes dataset using KNN with an accuracy of 94%.

## SKILLS

- PROGRAMMING SKILLS**  
Python, R, Java, C, C++, SQL
- MACHINE LEARNING TECHNICAL SKILLS**  
Time Series Analysis, Supervised and Unsupervised Learning, Ensemble Methods, Data Analytics, Image Processing, NLP, Big Data
- DATA TOOLS AND TECHNOLOGIES**  
Keras, PyTorch, TensorFlow, OpenCV, CUDA, Graph DB, MongoDB, Sci-kit learn, Pandas, SageMaker, Hadoop, Hive, Spark, HBase, BigQuery, Airflow, Amazon Redshift
- DATA ANALYTICS AND VISUALIZATION**  
Tableau, Adobe Analytics, Power BI, Plotly, Dash, Seaborn, Matplotlib

## CERTIFICATION

AWS Machine Learning Specialty (941/1000)

Jun 2020 - Jun 2023

TensorFlow Developer Certificate

May 2021 - May 2024