SHANTANU **HADAP**

Email: hada6576@stthomas.edu | Phone: 612-227-6051 | Address: 50 Cretin Ave S #105 St Paul MN

Profile

Experienced Machine Learning Engineer with demonstrated ability to deliver valuable insights via analytics, artificial intelligence, machine learning, and other advanced data-driven methods. Highly adept at extracting, analyzing, and interpreting large datasets using machine learning, deep learning, NLP, and other advanced algorithms. Strong diligence, and significant ability to work in a team environment, currently looking for a Data Scientist position with a forward-moving company.

Skills

- **Programming Languages**
  Python, Java, R, C#, C++
- **Machine Learning Algorithms**
  Univariate/Multivariate Regression, Decision Tree, Random Forest, XGBoost, SVM, KNN, K Means, PCA, Recommender Systems, Filtering, Gradient Descent
- **Deep Learning Algorithms**
  Neural Networks, CNN, RNN, LSTM, GRU, Auto Encoder, Object Detection, Segmentation
- **Data Mining/Extraction**
  NumPy, Pandas, NLTK, NLTK, SciPy, sklearn, Data extration, Feature selection, Feature extraction
- **Natural Language Processing (NLP)**
  NLTK, Spacy, Word2Vec, Glove, Word Embeddings, Gym, BERT, GPT, Tesseract, Dialogflow, Alexa Skills
- **Computer Vision**
  VGG16/19, Inception, ResNet, MRCNN, SegNet, Detecron2, Yolo, OpenCV
- **Forecasting**
  ARIMA, SARIMA, Prophet, LSTM
- **Database**
  SQL, Oracle, PostgreSQL, MySQL, MongoDB
- **Statistics**
  Exploratory data analysis, Hypothesis test, T-test, F-Test, ANOVA, Chi-squared test, Regression test, Significance and confidence interval
- **Data Visualization**
  Power BI, Tableau
- Cloud Services
  AWS, S3, AWS CLI, RDS, EB, Lambda functions, CloudFormation, Code Pipeline, SQS, SNS, CI/CD pipeline, SageMaker, DevOps, MLOps
- **Big Data**
  Hadoop, Spark, NiFi, Pig, Spark, Hive, Druid, Ozzi
- **Scripting**
  Linux script, AWS CLI, CloudFormation, Ansible
- **Tools and Libraries**
  TensorFlow, Keras, PyTorch, SageMaker, MATLAB, Jupyter Notebooks
- **Version Control**
  Git, GitLab
- **Agile Tools**
  Jira

Experience

Delta Airlines, MN — AI & ML Developer

01/2020 – 06/2020

- **Priority Code Classification and Forecasting** Delta employees get different types of free tickets for them, their family, and partner, these tickets are referred to as priority codes (S1-S4). The HR team wanted to analyze the priority codes to decide on policy changes.
- Performed data extraction, data cleaning, feature selection, and complex data preprocessing on the travel data collected from different multi-dimensional data sources using PySpark.
- Developed a forecasting model to predict future usage for the priority codes using SARIMA.
- Used XGBoost to develop a classification model to predict the priority code, based on collected travel data.
- **Customer complaints classification** Collaborated with the customer service team to developed and productionized an NLP solution to classify customer complaints from different text sources into related intents, this reduced waiting time, and resulted in improved and fast customer care service.
- Developed infrastructure, end to end delivery pipeline, and endpoints for deploying model as a service using AWS

SageMaker

- Developed an ETL pipeline that dramatically improved the speed of delivering data from HDFS to business reports using Apache Spark and Hive.
- Primary technologies and libraries used included Python, TensorFlow, Keras, NLP, SARIMA, SageMaker, S3, NLTK, XGBoost, GRU, LSTM, CloudFormation, Code pipeline, Hadoop, Apache, Spark, etc.

STELAR Research Lab, MN — Research Programmer, AI & ML

*12/2018 – Current (*Worked part-time when I was working with Delta)*

- U Buddy Developed an NLP backend service that handles voice requests using google Dialogflow API and cloud functions. This service enables handsfree usage of university portals like canvas, student account, library services. link
- Wheat detection Developed and productionized a model to detect wheat heads from outdoor images of wheat plants leveraging state of the art deep learning networks and libraries like Yolo, Detectron2, and OpenCV. The model was trained on more than 5000 images and achieved a map of 90.9 (0.5 IOU) link
- OCB Admission forecasting Created a forecasting model to predict the upcoming students for OCB MBA programs. Collaborated with the admission team to design real-time Power BI dashboards to help them make administrative and logistic decisions.
- UPDRS Score prediction Applied machine learning ensemble techniques for prediction of UPDRS score in Parkinson's disease link
- Open Images visual relations Developed a solution for detecting pairs of objects and relationships connecting that. The object detection model developed on the MRCNN network using Open Images dataset link
- Sentiment analysis Developed a multiclass sentiment classification model classify car recall tweets data extracted using GetOldTweetsV3 for the year 2018-2020. The project aimed to discover the sentiment triggered by car recall of different car manufacturers. link
- Journal OCR Developed a solution to digitalize Henry Thoreau's scanned journals using tesseract link
- Created a deployment pipeline for GiveMeFever and deployed it on AWS EB running Django stack.
- Major responsibilities for this role included data mining, feature extraction, feature selection, model tuning, creating pipelines, creating endpoints,
- Technologies and libraries used were AI/ML Algorithms, TensorFlow, Keras, SKLearn, NLP, CNN, NLTK, SARIMA, GPT, Power BI, and cloud technologies like S3, EB, SageMaker, CloudFormation, SNS, SQS.

University of St Thomas, MN — Graduate Research Assistant, Prof. Chi Lai

08/2019 – 01/2020

- The project aimed to analyze the toxicity level of water bodies using the images of fish embryos exposed to 14 different toxic chemicals, with different concentrations and exposure time. The morphological effect of the chemicals on 8 different organs of the fish embryo was also needed to be analyzed.
- Developed a deep learning CNN model to classify the images into different chemical groups using networks like VGG19, ResNet, Inception models with an f-score ranging from 0.88-0.95.
- Built a state of art a segmentation model to detect and mask different fish embryo organs like eye, bladder, notochord, yolk, capillary using segmentation networks like SegNet, Detectron2, MRCNN, FasterRCNN, Yolo with AP (IOU 0.5) ranging from 44-55.7
- Worked solo to develop an end to end solution, major responsibilities were data extraction, feature selection, image preprocessing, model tuning, and evaluation.
- Technologies and libraries used were Python, Keras, SKLearn, Detectron2, MRCNN, FasterRCNN, Yolo, MATLAB, and cloud technologies like S3 and EB.

SpiderLogic — Data Analyst / Software Developer

10/2016 - 08/2018

- Wipfli – Developed forecasting models to forecast daily, weekly, monthly operational, and sales performance. Analyzed data in revenue, sales, profit, opportunity, and risks from different dimensions by time, product, and account. Integrated custom visual reports based on requirements using Power BI.
- XPO DMS & eTrac - Developed and designed a software solution that facilitates the supply chain and logistics processes of XPO logistics inc. The system included an inventory management web app, separate mobile apps for drivers/installers and customers, and microservices to integrate those applications. The system is used to manage transportation, contract logistics, and supply chains for more than 50,000 XPO customers across North America,

including 69 Fortune 100. Collaborated with a team of 12 developers, major responsibility was requirement gathering and estimation, designing and developing code modules, unit testing, creating CI/CD pipelines.

- Pallex Nexus - Nexus is a software system that is used to streamline cross-partner communication and tractability, acting as a bridge between all domestic fright logistic and transport management systems across Europe. Enhanced the integrity and implemented new code functionalities for the existing system. Technology stack included .NET MVC, SQL, Azure, Bootstrap, Angular, etc.
- SERA – Responsible for rebuilding and redesigning the reimbursement management system using Azure PAAS stack, including Azure apps, MongoDB, Azure storage, SendGrid, notifications, and queues. Designed a streamlined deployment and delivery pipeline for the application.
- Followed Agile software development practices including sprint planning, weekly standups, iterations, retrospectives. Major technologies involved were Python Java, C#, SQL, PostgreSQL, MongoDB, Power BI, Web Services, Hibernate, AngularJS, JavaScript, Azure CI/CD, AWS/Azure Services, Jira, etc. Used Git as a version control tool.

Atos Syntel — Software Developer
08/2013 - 09/2016

- Procter & Gamble - Designed and developed inventory management and supply chain system for tracking and managing chemicals required for P&G products. Led a team of 5 developers to design and develop programs and processes to manage more than 100k chemicals across the distributed cost and profit centers. The technology stack was Java, SAP, Web Services, and Excel.
- CEP & DSMS - Implemented and deployed web-based contract execution, project recruitment, and resource platform for Atos. The platform enables Atos to manage more than 80k resources for worldwide IT project contracts. Collaborated with a team of 10 developers to implement the system in the .NET stack with the SQL database in the backend.
  Followed Agile software development practices including sprint planning, weekly standups, iterations, retrospectives. Major technologies involved were Java, C#, SQL, SSRS, JavaScript, Azure, SAP, VSTS, and TFS as a version control tool.

Education

2018 – 2020 — Master of Science in Data Science - University of St. Thomas, MN - GPA 4
Machine Learning, Artificial Intelligence, Data Analytics, Data Visualization, DevOps, Big Data

2009-2013 — Bachelor of Engineering in Electronics - University of Pune, India - GPA 3.8
Micro-controllers and Application, Artificial Intelligence, Embedded systems, Digital Image Processing, Operating Systems Architecture

Awards and Certifications

- MinneMUDAC 2019 Student Data Science Challenge – Top Performer
- FASTCON Prediction for Soybean Future prices - 2nd
- AWS Marketplace Dev Challenge Finalist
- Deep Learning Specialization (deeplearning.ai)
- Machine Learning (Stanford)
- Mathematics for Machine Learning Specialization – Linear, Multivariate, PCA (Imperial College London)
- Microsoft Certified Professional for Developing Microsoft Azure Cloud Solutions 70-532
- SAS certificate in Machine Learning

Research / Publications

- CIIS 2019 - Integrating and accessing University Information APIs using Natural Language Processing Tools
- Analysis of Fish Embryo Images using AI