# SUMITH REDDI BADDAM

sumith.reddy2@gmail.com |+1(408) 813-5534| https:/linkedin.com/in/sumith-reddy-baddam|https://github.com/SumithBaddam

## EDUCATION

| | | |
|---|---|---|
| Master of Science in Data Science | Indiana University Bloomington<br>*CGPA*: 3.95/4 | Dec 2020 |
| Master and Bachelor of Technology in Computer Science | IIIT Bangalore | June 2017 |

## PATENT AND RESEARCH PUBLICATIONS

- **Prediction of issues customers face in a software using unsupervised learning** Cisco Patent – 2020

Implemented Deep Neural Network model in TensorFlow which predicts the issues customers might face in a Cisco product post its release, helping developer teams fix them prior with an accuracy of 95% on Cisco's Next-Gen devices

- **Customer Success using Deep Learning** Advances in Economics and Business Vol. 6(6)

Built a prediction model for prioritizing the bugs identified during testing phases whether to be fixed fast or can wait. Unstructured bug attributes like descriptions, error log files along with 170 structured fields were used for building the system. It was implemented using LSTM and CNN in Keras and TensorFlow.

- **Intelligent defect creation system using Siamese CNN LSTM techniques** ICBAI, 2018

Implemented a duplicate bug detector that identifies whether a newly created bug is a duplicate of an existing bug in the Cisco Defect Tracking System and then retrieves all similar bugs from the database with an accuracy close to 90%

## PROFESSIONAL EXPERIENCE

**Data Scientist,** Alexa Shopping April 2021 – Present

Building Neural Language Models for Alexa shopping domain using NLP and recurrent neural networks.

**Software Development Engineer,** Amazon Web Services May 2020 – March 2021

Built an end-to-end large scale distributed machine learning feature for AWS CloudFormation to estimate resource provision time for customers. The pipeline consists of data extraction from S3 service, pre-processing using AWS lambda functions and the prediction models built and hosted on Amazon SageMaker.

**Data Scientist,** Cisco Systems Jan 2017 – Aug 2019

Built machine learning models to improve the quality of Cisco products and its internal workflow:

- Recommendation engine for identifying peer reviewers for testing on Cisco's code review platform using NLP.
- Keywords extraction and document classification of service request cases using unsupervised LDA modeling.
- Classification of Cisco products into various categories to help the sales teams improve their revenue generation.
- Identification of files that get impacted when set of files are committed to repository using Association Mining.
- Clustering the features of products based on the text data and summary fields with NLP and K-means clustering.
- Software upgrade recommendations to customers using random forest and data mining.

**Big Data Analyst**, Dataweave Software Pvt. Ltd., India

Built an automation engine that web scraps products data from various e-commerce websites and analyses customer reviews, classifies products into categories and clusters similar products into groups using computer vision and natural language processing to provide insights to customers. It was built to scale to 100 Million products concurrently using distributed systems.

## TECHNICAL SKILLS

- **Languages:** Python, R, R Shiny, Java, C, C++, Matlab, NodeJS, React
- **Platforms:** TensorFlow, Keras, OpenCV, Tableau, Scikit-learn, Flask, Django, AWS
- **Database:** SQL, AWS, Google cloud, MongoDB, Hadoop, Spark, NoSQL, JDBC, ZoDB

## KEY ACADEMIC PROJECTS

**NeuralCook – Image2Ingredients and cooking recommendation using Deep Learning:** (Demo)

- Deep learning application to identify ingredients from cooking dish images and recommend dishes to cook.
- End to end API leveraging Computer Vision and NLP using joint embedding space. Hosted on AWS.

**Human Robot Interaction using Natural Language Processing and Computer Vision:**

- Implemented 3-layer virtual assistant equipped with chat/dialogue bot, video, and speech analysis in Python.

**Implementation of Deep Neural Networks for Object Recognition in Python:**

- Visual categorization of objects using Convolution Neural Networks and Principal component analysis.

**Automated Essay Grading System:**

- Feature extraction on text data using POS Tagger, Word2Vec and modeling with NLP and ensemble learning.
- Implemented Association rule mining, classification, clustering, and statistical analysis to extract insights.

**Bayesian Belief Networks for Restaurant violations predictions:**

- Neural network and Bayesian belief networks for predicting violations of a restaurant using Yelp dataset.