CHAITHRA K. C

315-728-0657 | [ckoppara@syr.edu](mailto:ckoppara@syr.edu) | [www.linkedin.com/in/chaithra-kc](http://www.linkedin.com/in/chaithra-kc)| [github.com/chaithrakc](file:///G:\Other%20computers\My%20Laptop\chaithrakc.github.io\resume\github.com\chaithrakc) | [www.kaggle.com/chaithrakc](https://www.kaggle.com/chaithrakc)

# EDUCATION

## SYRACUSE UNIVERSITY, SCHOOL OF INFORMATION STUDIES Syracuse, New York

## Master of Science Applied Data Science May 2023

Coursework: Introduction to Data Science, Applied Machine Learning, Applied Deep Learning, Natural Language Processing, Business Analytics, and Data Analysis & Decision-Making

**JAIN UNIVERSITY Bangalore, India**

## Bachelor of Engineering Computer Science Jun 2015

Relevant coursework: Principles of Programming Languages, Java & J2EE, Data Structures, Algorithms, Database Management Systems, Data Mining, Business Intelligence, Design Patterns, Unix System Programming, Linux Internals

# TECHNICAL SKILLS

**Key Skills:** Machine Learning, Natural Language Processing, Deep Learning, Data Cleaning, Analysis and Visualization, Hypothesis Testing, Probability and Statistics

**Programming Languages**: Python, R, Java, SQL, PL/SQL, NoSQL

**Python Libraries**: PySpark, Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn, NLTK, Spacy, Keras, TensorFlow, Gensim, Beautiful Soup, PDFMiner

**Databases**: MySQL, Oracle DB, Mongo DB

**Tools & Utilities:** Databricks, Apache Spark, Tableau, Advanced MS Excel, MS Access, Perforce, Git, Maven

**IDEs:** Google Colab, Jupyter Notebook, PyCharm, R-Studio, IntelliJ

# EXPERIENCE

## Data Science Intern, RSG Media, New York May 2022 - Aug 2022

## Developed a data pipeline for model input - modify

## IMDb and Viacom data ingestion - modify

## Entity Resolution between Viacom’s program inventory & Gracenote data, and IMDb & TMDB implemented using K-Nearest Neighbor ML model, TF-IDF word vectorization, and NLP pre-processing techniques

## Created a prediction report using Microsoft Excel to compare predictions across multiple models

## Senior Software Engineer, Envestnet Yodlee, Bangalore, India Mar 2018 - Sep 2020

* Developed data quality tools in Java to validate holding and transaction data veracity and extrapolate missing financial details based on patterns observed with other users
* Modified sequential to a multi-threaded application to process millions of investment accounts and transactions per day
* Implemented Netflix-Ribbon load balancer to achieve fault tolerance and distribution of application traffic
* Implemented POC for determining suitable database for the project among OLTP (Oracle DB), Mongo DB, and Apache Kudu based on query performance
* Provided Mongo DB support for the project that includes designing Collections, designing Mongo documents, Indexing Collections, and Configuring Spring repositories to query from the application

## Software Engineer, Envestnet Yodlee, Bangalore, India Jul 2015 - Feb 2018

* Built web crawlers using Java Selenium APIs to aggregate users' financial data and personal information from banking websites
* Created PDF parser for retrieving data from bank statements using Python PDFMiner text extraction tool
* Implemented regex validator using Java regular expressions for validating PII
* Performed peer code review and mentored a few new joiners

# ACADEMIC PROJECTS

**News Summarization**  **Mar 2022 – May 2022**

* Project's goal is to generate coherent summaries of news stories using different Deep Learning models - T5 Transformer, Encoder & Decoder with BiLSTM, and NLP word embedding

**Sentiment Analysis of COVID News Articles**  **Mar 2022 – May 2022**

* Predicting the sentiment polarity (Negative, Positive, and Neutral) of COVID news articles obtained through IEEE data port using Naïve Bayes classifier and BiLSTM model
* Bag of Words (BOW) model, BERT word embedding, and XLNet embedding are used to transform text into vectors