CHAITHRA K . C

315-728-0657 | [ckoppara@syr.edu](mailto:ckoppara@syr.edu) | [LinkedIn](http://www.linkedin.com/in/chaithra-kc) | [GitHub](https://github.com/chaithrakc) | [Kaggle](https://www.kaggle.com/chaithrakc)

# EDUCATION

## MASTER OF SCIENCE APPLIED DATA SCIENCE May 2023

**Syracuse University - School of Information Studies Syracuse, New York** Relevant coursework: Intro to Data Science (R-Studio), Business Analytics (Statistics, Excel, Access, Tableau, MS Power BI, Google Analytics), and Data Analysis & Decision Making (Statistics using Excel), Machine Learning, Deep Learning, and Natural Language Processing.

CGPA: 4.0/4.0

## BACHELOR OF ENGINEERING COMPUTER SCIENCE Jun 2015

**Jain University Bangalore, India**

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 CGPA: 9.56/10

# 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, HTML, CSS

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

**Databases**- MySQL, Mongo DB

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

**IDE Tools**- Google Colab, PyCharm, Jupyter Notebook, R-Studio, Intellij

# EXPERIENCE

## Data Science Intern, [RSG Media](https://www.rsgmedia.com/), New York May 2022 - Aug 2022

## Developed a data pipeline for model input

## IMDb and Viacom data ingestion

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

## Automating the ETL process for matching movie titles

## Create a prediction report to compare model predictions across multiple models

## Skills - NLP, Unsupervised ML, Databricks, Apache Spark, SQL, Python, Amazon S3

## Graduate Teaching Assistant, Information Technologies, Syracuse University, New York Aug 2021 - Present

* Facilitate weekly labs for 30-40 undergraduate students
* Enhanced Linux, AWS Cloud Computing, and Security labs to reflect the concepts taught in the lecture
* Other duties included holding weekly office hours, grading assignments & projects, developing exams, proctoring exams, and so on

## Senior Software Engineer, [Envestnet Yodlee](https://www.yodlee.com/), 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](https://www.yodlee.com/), 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** ([GitHub Link](https://github.com/chaithrakc/News-Summarization)) **Mar 2022 - Present**

* Project's goal is to use different Deep Learning techniques - T5 Transformer, Encoder & Decoder with BiLSTM models, and NLP to generate coherent summaries – to generate brief descriptions of news stories

**Hotel Booking Cancellation Prediction** ([GitHub Link](https://github.com/chaithrakc/hotel-booking-cancellation-prediction)) **Nov 2021 - Present**

* Project's goal is to develop data-driven solutions utilizing machine learning techniques to help hotels in cities and resorts reduce cancellations while improving revenue and profits.