ANUSHA CHIMMILI

Data Science professional with expertise in building end-to-end ML solutions

anusha.chimmili.ofc@gmail.com | GitHub | LinkedIn | (636) 409-4852 | St. Louis, MO

EDUCATION

University of Central Missouri, Missouri — Aug 2022 - Dec 2023

Master of Science, Big Data Analytics & Information Technology — GPA - 3.9/4.0

Coursework - Business Applications of Machine and Deep Learning, Business Intelligence and Analytics, Big Data Solutions for Business, Information Visualization, Big Data Architecture, Data Resource Management

Indian Institute of Technology Bombay, India — Jul 2016 - June 2018

Master of Technology, Biomedical Engineering — GPA - 9.5/10.0

Coursework - Applied Statistics, Fundamentals of Image Processing, Algorithms for Medical Image Processing

Osmania University, India — Sep 2011 - May 2015

Bachelor of Technology, Biotechnology — GPA - 8.4/10.0

Coursework - Introduction to Biostatistics, Bioinformatics, Python Programming

PROFESSIONAL EXPERIENCE

Aureus Tech Systems — Kansas City, MO

Data Science Intern — Jan 2023 - May 2023

- Categorized and extracted 15 types of Personally Identifiable Information (PII) using SpaCy, BERT, NLTK, regular expression tools, from text documents to provide a PII tool with enhanced data privacy and compliance.
- Improved the accuracy of identifying names, addresses, and Social Security Numbers (SSNs) by 15% through rigorous quality testing and post-processing methods.
- Implemented the optimized model in production using Databricks to analyze ~30k text files, streamlining the PII redaction process.

Target Corporation — Bangalore, India

Senior Data Scientist — Oct 2021 - Jul 2022

- Developed and implemented Generalized Additive Mixed Models (GAMM) to accurately forecast demand for ~300k items on Target's Digital Platforms, resulting in a 20% reduction in inventory replenishment time and better product placement.
- Consolidated and streamlined data from 6 different database systems to provide comprehensive visibility of item eligibility across all the fulfillment types on digital platform, increasing overall efficiency by 15%.
- Successfully migrated 7 workflows from legacy systems to the Spark3 ecosystem, ensuring enhanced data security measures and reducing potential risks by 30%.

Data Scientist — Jul 2020 - Oct 2021

- Developed a comprehensive validation framework by creating KPI dashboards at different grains to track the accuracy and bias of forecasted sales for ~100k items, which helped in model improvements across various item assortments.
- Optimized model accuracy by tuning hyperparameters for trend and seasonality components, resulting in an average increase of 10-15% in forecast accuracy for slow selling item departments.
- Scaled the forecasting pipeline from 7 to 60 departments out of the total 110 under Target's inventory list, while simultaneously reducing the runtime by 2 hours, allowing robust weekly forecast generation.

Data Analyst — Dec 2019 - June 2020

- Streamlined the reporting dashboards for labor planning during peak season by collaborating with Sales & Operational Planning business teams, resulting in a 50% reduction in time spent on report generation and analysis.
- Improved labor forecasts at fulfilment centers by generating future order profiles based on the stratified sampling of past order profiles, resulting in a decrease of overstaffing by 20% during peak season.
- Automated the generation of 21 reports from different databases critical to labor planning in distribution centers, eliminating the need for manual updating and reducing data errors by 15%.

EXL Inductis Analytics — Bangalore, India

Data Analytics Consultant — Jul 2018 - Oct 2019

- Used 110 consumer specific attributes from Experian data to identify potential customers for targeted marketing with XGBoost, resulting in a 15% increase in sales.
- Collaborated with cross-functional teams to develop and implement marketing strategies based on data-driven insights, resulting in a 10% increase in customer acquisition.
- Implemented sentimental analysis on customer reviews, providing actionable insights to improve product offerings and customer satisfaction levels.

ACADEMIC PROJECTS

Caption Generation — Nov 2023

- Developed a web application that generates captions for user provided images or image URLs using fine-tuned hugging face models - BLIP and GIT-Base-COCO.
- Created the frontend interfaces with HTML, CSS, JavaScript using Flask framework and customized them for different user tiers.
- Designed the backend database system using Entity Relation Model for storing images, image tags and model generated captions, to efficiently retrieve captions using date, tag or caption content filters.

Dog Breed Identification — Aug 2022

- Creating a classifier that can determine dog's breed based on an image out of the total 120 breeds of dogs presented in the dataset.
- Developed custom data pipeline for image pre-processing and augmentation using PyTorch and OpenCV.
- Pre-trained classification models such as VGG16 and Inception v3 were used to classify the dog breed using PyTorch.

Entity Extraction from Research Paper Abstracts — Feb 2021

- 300+ scientific paper abstracts were scrapped using BeautifulSoup4 library, out of which 200 were used for labeling process.
- Data is annotated using Brat Rapid Annotation Tool and then later converted from BRAT standoff format to Conference on Natural Language Learning (CoNLL) format.
- Implemented a named entity recognition (NER) system to extract author name, year of publication and keywords from the abstracts of the scientific research papers.

Character and Face Recognition — Oct 2019

- Built a Bayesian Classifier using 200 training images of each character to categorize the 100 test characters using multi-dimensional Gaussian distribution and Maximum Likelihood techniques in MATLAB.
- Built a polynomial regressor and learn its weights by adopting a K-fold cross-validation strategy.
- Face reconstruction and recognition performed for 200 Images of 40 Individuals using KL Transform (Principal component Analysis) and Fisher linear Discriminant for classifying face images to their correct class.

Master's Thesis Project — June 2018

- Conducted a study on the effect of fatigue on cognition in a repeated measures design to collect the measures of subjective and objective tiredness.
- Performed Exploratory Data Analysis (EDA) on the collected data in MATLAB and tested the hypothesis using 2-way ANOVA as well as tested the statistical power of the effect of fatigue.

TECHNICAL SKILLS

Machine Learning Techniques: Generalized linear models, Tree-based models, Ensemble methods, Time series forecasting, Neural Networks, Natural Language Processing models, Exploratory Data Analysis, Dimensionality Reduction, Clustering models.

Statistical Techniques: Hypothesis Testing, t-Tests, ANOVA, Descriptive and Inferential Statistics

Programming Languages: Python, SQL, R, PySpark

Python Libraries: Scikit-Learn, Prophet, Hugging Face, BERT, NLTK, SpaCy, Keras, PyTorch, TensorFlow

Technologies: Big Data tools - Hive, Hadoop, Spark; CI/CD tools - Docker, Jenkins, Maven; Cloud - AWS EC2, S3

HONORS & ACTIVITIES

- **Chief Data Science and Analytics Officer Award** 2021: Digital Demand Forecasting, Target.
- Presented thesis project as a poster at **48th Annual meeting of Society of Neuroscience 2018** in San Diego, California.
- Granted a **Graduate Assistantship** with a full semester scholarship during my master's program at the University of Central Missouri.
- **Work Eligibility:** Extended eligibility to work in the U.S.; will require visa sponsorship for long-term employment.