

# Bhumika Srinivas

[Email](#) | [LinkedIn](#) | [GitHub](#) | [Portfolio](#) | +1 (415) 670-0906 | San Francisco, CA

---

## EDUCATION

### Master of Science in Data Science

University of San Francisco

Relevant Coursework: Machine Learning, Deep Learning, Data Acquisition, Statistics, Distributed Computing

*Jul 2023 – Jun 2024*

*San Francisco, CA*

### Bachelor of Science in Data Science

University of San Francisco

*Aug 2019 – Dec 2022*

*San Francisco, CA*

---

## PROFESSIONAL EXPERIENCE

### Data Scientist

*Oct 2023 - Jun 2024*

*UCSF Medical Center - Radiation Oncology*

*San Francisco, CA*

- Optimized over 12 million words of cancer patient clinical notes and EHR data through meticulous preprocessing, achieving a remarkable 25% text reduction for enhanced data efficiency.
- Engineered two advanced pipelines for hierarchical segmentation using encoder-driven transformer models, resulting in an 8% accuracy improvement with XGBoost over Logistic Regression and Random Forest models, enhancing early disease detection for over 5000 patients and optimizing their treatment plans as an intern.
- Strategically segmenting lengthy documents to overcome token limits in models resulted in a 30% reduction in computational overhead, thereby enhancing processing efficiency and lowering GPU utilization while maintaining contextual coherence.
- AAPM abstract submission: Explored Baseline MPT-7B and fine-tuned clinicalBERT; revealed LLMs don't inherently enhance EHR notes' predictive accuracy, advocating for fine-tuned models in diverse cancer patient notes.

### Data Scientist

*Feb 2023 – Jun 2023*

*ReshaMandi*

*Bangalore, India*

- Enhanced silk farm profits by 37% (\$8,000 monthly) with an Amazon SageMaker-based machine learning model, deploying a serverless inference pipeline and integrating it into a Kafka consumer pipeline for optimal cocoon procurement decision-making.
- Implemented a Tableau project in collaboration with the engineering team, utilizing Snowflake SQL queries to create a dynamic dashboard that resulted in a 25% reduction in turnaround time (TAT) for logistics processes for the 17-member SCM team.
- Optimized Neo4j Cypher queries for existing knowledge graphs, achieving a 40% performance boost and saving 21 hours weekly, thereby expediting information retrieval, enhancing data accessibility, and improving decision-making capabilities as an intern.

### Data Analyst

*Jun 2022 – Aug 2023*

*Object Edge*

*Walnut Creek, CA*

- Increased sales by 40% by leveraging insights via R-based RFM analysis, evaluating customer satisfaction and targeting specific demographics based on order dates, monetary value, and product quantity, managing a dataset of over 11 million records.
- Enhanced website order checkout efficiency by 55% through Market Basket Analysis using Python, optimizing product placement and encouraging targeted sales combinations as an intern.
- Designed and presented a Tableau dashboard for Zenni Optical, visualizing 5 key business KPIs, leading to actionable insights into Gross Merchandise Value and Order Revenue.

### Data Analyst

*Mar 2022 – Dec 2022*

*University of San Francisco*

*San Francisco, CA*

- Led multiple data visualization research projects employing Qualtrics, R, Python, Excel.
  - Presented findings to the University of Nottingham team which involved conveying insights, methodology, and implications, fostering collaboration and knowledge exchange within the academic community.
  - **Published** a co-authored research paper on treemap literacy, accepted by [Eurographics Education 2023 Papers](#).
- 

## PROJECTS

- **Deep Learning - [Image Segmentation of Microscopy Images](#)**
    - Placed first in this [Kaggle Competition](#). Customized U-Net neural network using TensorFlow, incorporating strategic layer variations to achieve unparalleled accuracy in predicting biomedical microscopy images of cells.
  - **Distributed Data Systems with Spotify Web API**- Apache Airflow, MongoDB, SparkSQL, and Google Composer
    - Developed and deployed an [automated data pipeline using Airflow on Google Composer](#) to create a localized social media platform for small friend groups, enabling users to connect their Spotify accounts, share listening history, and conduct analytics.
- 

## TECHNICAL SKILLS

Python (Matplotlib, SciPy, Plotly, Scikit-Learn), Java, SQL (Postgres, Snowflake), Git, GitHub, PyTorch, TensorFlow, BeautifulSoup, RegEx, Time Series Forecasting, A/B Testing, PySpark (Databricks), Airflow, NoSQL (MongoDB), R (tidyverse, ggplot2), LaTeX, GCP, AWS

---

**CERTIFICATES** Data Institute Award (Outstanding Graduating Senior in Data Science), **AWS Certified Cloud Practitioner**