# Sanidhya Mangal

615-955-8605 mangalsanidhya19@gmail.com /sanidhyamangal /sanidhyamangal

## Education

**Vanderbilt University** *(Nashville,TN)*, MS in Computer Science; *3.94/4.0* **December 2022**

**Medi-Caps University** *(Indore,India)*, B.Tech in Computer Science; *8.41/10.0* **May 2020**

## Technical Skills

**Development Tools**: Python (Pandas, Scikit-Learn, Numpy, Tensorflow, Keras, Pytorch, Django, Flask), Bash, Scala
**Analysis Tools**: SQL (Presto, MySQL, Oracle, DynamoDB, ORM), MS Excel, MS PowerBI, Tableau, Hadoop
**Deployment Tools**: Docker, Kubernetes, GIT, AirFlow, AWS (Lambda, EC2, EKS, ECS, RDS, S3, Sagemaker)

## Experience

**Asurion** **May 2022 – August 2022**

*Data Science Intern* *Nashville, TN*

- Designed SVM and Decision Tree models stacked on TF-IDF to analyze real-time call transcriptions, influencing expert behavior and driving upsell, resulting in an estimated 6% improvement in sales, equivalent to $1.2 million.
- Assembled DS life-cycle: ideation, opportunity sizing, modeling, deployment, and exposure in A/B testing.
- Engaged with stakeholders to identify & define business & analytical needs; translated insights into business outcomes.
- Experimented with feature engineering & BERT variants to improve classification model robustness on a small dataset.
- Achieved 88% of overall recall with a mean AUC of 0.81 along with meaningful insights like correlation coefficients.

**Maize Zhou Lab** **July 2021 – May 2022**

*Research Assistant* *Nashville, TN*

- Contributed to research projects by developing a toolkit for state-of-the-art ML models to facilitate genome filtering on long and short reads, leading to multiple publications; improvement of overall F-1 score by 20% from predecessors.
- Translated research problem into end-to-end MLOps pipeline, including data processing, modeling, and evaluation.

**Engineerbabu** **June 2020 – June 2021**

*Machine Learning Engineer* *Indore, India*

- Desinged a CNN based tool perform progronis of lung and colon cancer with overall AUC of 0.92 and 91% precision.
- Supervised a team of six that reduced inference time by 30% for machine learning models by improving the ML pipeline.
- Deployed linear/GLM models in production for edge AI using API, enabling interpretation & generation of insights.
- Optimized decision-making process by 12% through EDA, including ETL, hypothesis testing, and statistical analysis.

**Greater Kailash Hospitals** **January 2020 – April 2020**

*Machine Learning Engineer Intern* *Indore, India*

- Developed a web app to provide a second opinion to physicians in diagnosing X-rays for Pneumonia with 94% precision.
- Improved baseline performance by 20% through bagging and ensemble methods, including Random Forest.

## Projects

**Self-SupervisedGAN** | *PyTorch, GAN, Self-Supervised Learning, Computer Vision, Generative Modelling*

- Generated high-fidelity images and improved VanillaGAN's performance using self-supervised pre-training.
- Added self-supervision to prevent forgetful discriminator which aids in better convergence and prevents mode collapse.

**Interpretable-Bert** | *PyTorch, Transformers, Interpretability, Named Entity Recognition, Classification, NLP*

- Designed probes to leverage pre-trained BERT representations to perform named entity recognition on the input text.
- Analyzed contextual representations to examine how pre-training task affects the linguistic knowledge in transformers.

**Date Matcher** | *Supervised Machine Learning, Logistic Regression, SVM, Deep Neural Networks*

- Trained a model to predict match and date probability based on personal attributes, achieving an F1 score of 0.98.
- Optimized model performance by using PCA, resulting in 20% reduction in dimensions and complexity of models.

**AutoMLify** | *Machine Learning, Python, JAX, Gaussian Process, Linear Regression, KNN, SVM, Decision Tree, XGBoost*

- Implemented library for autotuning ML algos: classifiers & regressors with bagging & boosting for structured data.
- Extended the library to support neural networks and optimized training time by 10% through parallelization.

**Color Compressor** | *Unsupervised Learning, K-Means, Clustering*

- Implemented K-means clustering to compress images, reducing color space from 16 million to 16 retaining image quality.

## Publications


- Yunfei Hu, **Sanidhya Mangal**, Lu Zhang, and Xin Zhou. "Automated filtering of genome-wide large deletions through an ensemble deep learning framework." Methods (2022).
- **Sanidhya Mangal**, Aanchal Chaurasia, and Ayush Khajanchi. "Convolution neural networks for diagnosing colon and lung cancer histopathological images." arXiv preprint arXiv:2009.03878 (2020).
- **Sanidhya Mangal**, Poorva Joshi, and Rahul Modak. "LSTM vs. GRU vs. Bidirectional RNN for script generation." arXiv preprint arXiv:1908.04332 (2019).