Sandeep Gunda

508 Main Street, Atlanta, GA, 30324, (404) 509-9301
sgunda1@student.gsu.edu, Linkedin.com/in/SandeepGunda1, Github.com/SandeepGunda

PROFESSIONAL SUMMARY

Data Science Graduate with internship at CNN TV Research and experience in data science projects, media research and analytics.

SKILLS

Programming languages:	R, Python, SAS, C++, Java, MATLAB
Data Science Packages:	Numpy, Pandas, Matplotlib, Plotly, SciPy, GGPlot2, Tidyverse, dplyr, Beautiful Soup
Data Analytics Frameworks:	TensorFlow, Keras, Scikit-Learn, OpenCV, NLP, NLTK
Data Science Techniques:	Linear / Logistic Regression, Decision Trees, Clustering Algorithms, SVM
Relevant Coursework:	Statistics, Econometrics, Machine Learning, Text Analytics, Image Analytics, Deep Learning
Data Visualization and Databases:	Tableau 10.5, Microsoft Power BI, NoSQL, Mongo DB, Neo4j, SQL
Cloud Platforms and Big Data:	Hadoop, Spark, Microsoft Azure, Amazon Web Services, Google Analytics
Development Tools:	Git, Eclipse, Anaconda, R Studio, Jupyter Notebook
MOOC(Coursera):	Applied Text Mining, Google Analytics for Beginners, Regression Models

EDUCATION

J. Mack Robinson School of Business, Georgia State University   **Atlanta, GA, USA**
Master of Science in Data Analytics, 3.9/4.0 GPA   *Dec 2018*
Bharati Vidyapeeth's College of Engineering, Guru Gobind Singh Indraprastha University   **New Delhi, India**
Bachelor of Technology, Electronics and Communication Engineering   *Jun 2017*

PROJECTS

Tree Log Identification (Georgia Pacific) – OpenCV, KL Divergence, TensorFlow, Deep Learning   *Aug 2018 – Dec 2018*
- Developed an object detection deep learning model to identify and count the number of tree logs in every truck image from the Georgia Pacific warehouse.
- Made an object segmentation tool to identify, segment and manually label the tree logs from the images to develop a labelled dataset using OpenCV and KL Divergence in python.
- Defined a deep learning model using U-Net architecture to predict the number of tree logs in over 11000 images with 78% accuracy.

Anti – Money Laundering Project (SunTrust Bank) – Python, Scikit-learn   *Jan 2018 – Apr 2018*
- Worked with the largest regional bank in the South-East USA which spends a considerable amount of time and resources investigating 30k+ suspicious money laundering alerts per month to develop a model which predicts the seriousness of the alerts.
- Led a team of 4 and tackled issues such as peer – profiling, unavailable data and false positives by feature generation and selection.
- Tested the data with Logistic Regression, Random Forest, SVC and ADA Boost models to predict the quality of alert (High or Low) using precision and recall parameters.
- Delivered the ADA Boost model with a precision of 38.50% and recall of 78.48% for its high recall prediction.

Trip Duration Prediction – Python, Scikit-learn, Keras   *Jan 2018 – Apr 2018*
- Designed a Machine Learning model to predict the taxi trip duration based on the pickup and drop-off coordinates.
- Used K Means clustering to determine the various routes across the city and predicted the trip durations using Ridge regression, Lasso Regression, Random Forests and Neural Network and improved the predictions using grid search techniques.
- Random Forest Regressor provided a prediction with adjusted R square of 88%, which can have a significant impact for cab companies to determine their pricing.

PROFESSIONAL EXPERIENCE

CNN (Atlanta, GA)   *Jun 2018 – Dec 2018*
Intern – TV Research (CNN Research and Analytics)
- Collect and analyze multiplatform television data across various sources, such as CNN linear TV data, VOD, SVOD, and digital.
- Aid Research Analysts in the production of daily and regular reporting needs for the CNN brand as well as competitive networks
- Extract and analyze Nielsen audience delivery data using a variety of research software (Ex: Nielsen NPower, Arianna, Star Media)
- Assist team with ongoing projects, data visualizations, client presentations (in Microsoft Excel, Microsoft PowerPoint).
- Assist in the development of a machine learning model to forecast the delivery of HLN based on historical data and other parameters.

J. Mack Robinson College of Business, Georgia State University (Atlanta, GA)   *Aug 2017 – Apr 2018*
Graduate Research Assistant (Computer Information Systems)
- Investigated the influence of deployment of Electronic Medical Records (EMRs) on medical care with an aim to find if cloud-based EMRs lead to an inferior quality of care than their on-premise counterparts.
- Web-scrapped EMR data in R using Httr, Rvest, and Tidyverse packages and linked it to the Federal Meaningful Use program data.
- Analyzed the relationship between crowdfunding and productivity by linking data from multiple sources to determine the influence of income of a YouTube-r from the former on the success of his YouTube channel.
- Developed a web crawler using Beautiful Soup and Urllib packages in Python to scrape and build datasets.

National Institute for Smart Government (NISG) (New Delhi, India)   *Jun 2016 – May 2017*
Intern, Corporate Data Management (CDM) Project
- Worked on a project to digitize and analyze Ministry of Corporate Affairs' corporate data obtained over the years.
- Made metadata of e – forms then extracted the data into Excel, summarized their characteristics and visualized using Tableau.