**VINITH VIRUPAKSHAPPA ANGADI**

Los Angeles, CA [Open to relocation] | angadivinith@gmail.com | (213)-929-9860 | https://[www.linkedin.com/in/vinith-angadi](http://www.linkedin.com/in/vinith-angadi) | <https://github.com/VinithAngadi>

**WORK EXPERIENCE [2+ years as Data Scientist]**

**Data Scientist | Meta Jan 2023 – Present**

* **Building Migration Optimization**
* Worked with non-technical stakeholders to build an analytical solution that gives insights and helps reduce real estate use by Meta.
* Built an **optimization model** to reorganize employee work groups into older buildings and formulated **objective functions** to incorporate constraints on meeting room demand and building capacity.
* The analytics pipeline helped stakeholders make informed decisions and resulted in **18%** **reduction** of real estate use by Meta across Menlo Park and San Francisco sites.
* **Space Captain Prediction**
* Utilized my skills to collaborate with stakeholders and automate the process of assigning Meta new hire and internal transfer employees to space captains responsible for desk allotment on a global scale.
* Performed feature selection with insights from stakeholders and statistical feature importance methods. Built and deployed a **Random Forest Classifier** to classify employees to space captains with **0.95 F1 score**. Scheduled periodic model prediction workflow using **Airflow**.
* Reduced assignment to space captain latency from **days to minutes**.
* **Data querying and visualization**
* Used **Presto SQL** to query data from **Hive** tables. Visualized data on dashboards using **Tableau**. Used **Airflow** to schedule execution of workflows.

**Engineering Data Scientist | Snap Inc. Aug 2021 – Nov 2022**

* **Re-define Casual and Power Users**
* Revised the approach for identifying Casual Users and Power Users, transitioning from **quantile regression** to clustering methods. Standardized the identification process, facilitating targeted segmentation of users for **A/B tests**.
* Developed and implemented **K-Means** model using user engagement features and scheduled periodic clustering of users using **Airflow**.
* Conducted an **A/B test** to vary ad load in casual andpower user clusters which resulted in **3% increase** in revenue from story ads.
* **Sponsored Lens Carousal Ranking**
* Transformed the sponsored lens placement strategy within the lens carousel by transitioning from static method to dynamic ranking, with the primary goal of enhancing user engagement.
* Implemented the incorporation of **Estimated Organic Value (EOV)** of users into the ad ranking model and conducted **A/B Test** to measure change in user engagement of sponsored lens.
* Concluded in **10% increase** in sponsored lens engagement.
* **Data querying and visualization**
* Utilized **Google BigQuery** to query data from cloud. Visualized data on dashboards using **Looker**. Used **Airflow** to schedule execution of workflows.

**TECHNICAL SKILLS**

**Programming Languages:** Python, SQL, Java, R, C#, Visual Basic.NET

**Packages: Python -** Pandas, NumPy, PySpark, sklearn, seaborn, matplotlib, OpenCV, flask; **R -** tidyverse, dplyr, ggplot2, forecast

**Databases:** Apache Hive, Presto, MSSQL Server, PostgreSQL, AccessDB, Google Firebase, Metabase, RedShift

**Software tools:** Git, GCP, AWS, Airflow**,** Docker, Tableau, Looker, Android Studio, SAS Viya, SAS JMP, R Studio

**EDUCATION**

**Master of Science in Analytics,** University of Southern California , USA **Jan 2020 - Dec 2021**

**Bachelor of Engineering in Computer Science,** KLE Technological University, India **Aug 2015 - Jun 2019**

**PROJECTS**

**NYC Property Tax Record Fraud Detection Jan 2021 – Feb 2021**

* Conducted a comprehensive analysis of Property Valuation and Assessment Data from NYC real estate data and generate fraud scores to identify top fraud properties.
* Visualized and imputed missing values in data, defined new metrics to aid in detecting anomalies in properties.
* **Utilized** **PCA** for dimensionality reduction and computed fraud scores using **z-score** and **autoencoders** to analyze and rank top fraud valuated properties.

**Job salary prediction and analysis Oct 2020 – Nov 2020**

* **Web scraped** and preprocessed job descriptions and salary data from job sites like LinkedIn, Indeed, Dice.com. Generated **TF-IDF** vectors out of job descriptions and used topic modelling to generate multiple features.
* Compared multiple regression models with their statistical significance to predict salary given a job description and location. Used **Gradient Boosted Regressor** with an **MAPE** of **13% (+/- $9,112)**.

**AWARDS AND LEADERSHIP ROLES**

* VGSA Senator, ISE Department, USC Viterbi School of Engineering **Jan 2020 – Apr 2021**
* Secured first prize among 100 teams in Developthon-2017, conducted by Sandbox Startups **Jan 2017**