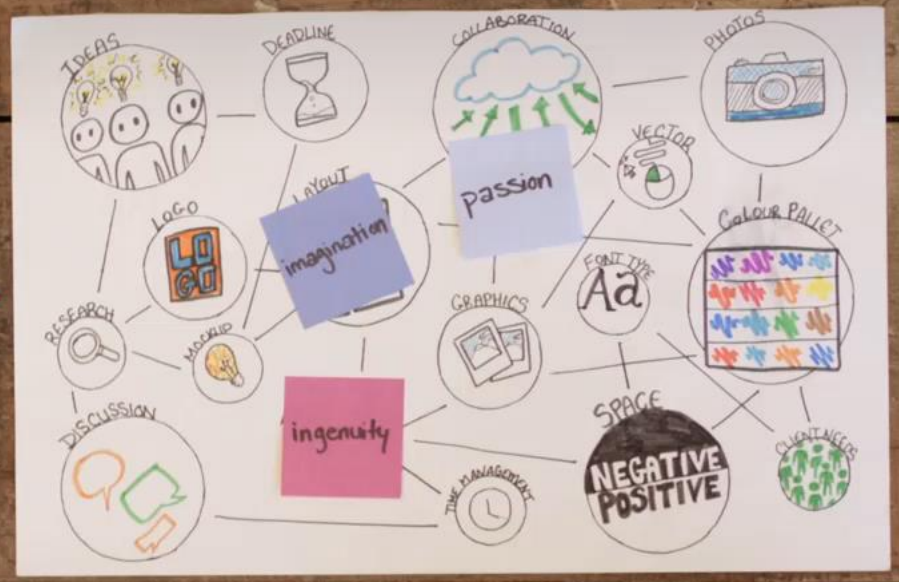
WEEK #8 MINI PROJECT



- Introduction
- Data Exploration and Preparation
- Model Building and Evaluation
- Results Interpretation and Conclusion
- Project Presentation
- Conclusion

INTRODUCTION

GOALS

- To prepare you for the real-world machine learning project.
- To practice on end-to-end model development workflow.
- To prepare you for the capstone project. (Treat this as the mini capstone project!)

EXPECTED OUTCOMES

- At the end of today, you should have created
  - An end-to-end machine learning development project with
    - EDA
    - Model building and evaluation
  - Presentation

DATA SCIENCE PROJECT OVERVIEW

TYPES OF PROJECT

- Regression analysis
- Classification problem
- Clustering problem
- Recommender system
- Time Series Forecasting

DATA

- Kaggle
- UCI
- Google

DATA EXPLORATION AND PREPARATION

- Perform a thorough EDA, including visualizations and statistics, to understand the distribution, correlation and missing values of data.
- Clean and preprocess the data effectively by handling missing values, outliers and other issues.
- Identify and justify the features used in the model and explain why they are relevant

MODEL BUILDING AND EVALUATION

- Create multiple models including base models
- Determine suitable models with appropriate metrics
  - Accuracy, precision, recall or F1-score
  - RMSE
- Evaluate models using k-fold
- Perform hyperparameter tuning
  - Grid search

PROJECT PRESENTATION

PRESENTATION

- Your slides should contain
  - Problem statement
  - Data source: Where you get your data?
  - EDA
    - Interesting insights that you can get from your data
    - Hypothesis drawn from the data
  - Model building and evaluation
    - Model development process
    - Justify why you choose the final model
  - Conclusion
    - Conclusions that can be drawn from the data and model

PRESENTATION

- Share me your code via:
  - GitHub (highly recommended)
  - Cloud drive
  - Share me the links here

- You're given 10 mins to present the slide.

CONCLUSION

CONCLUSION

- 9.30am – 12.00pm: Find dataset, develop model
- 12.00pm – 1.00pm: Lunch
- 1.00pm – 2.00pm: Develop model
- 2.00pm – 2.30pm: Wrap up, create slides
- 2.30pm – 4.00pm: Presentation, review model development process, code

CONCLUSION

- ENJOY!