



# Home Loan Default Risk Prediction

(Team - Caesar)

Team Member 1 - [Dhruv Awasthi](#) (Point of Contact)

Team Member 2 - Ashok Senapati



# Key Features

- Command line interface
- Built as a pipeline
- Modular approach
- Abstract components
- Logging everything
- Exception handling
- Organized directory structure
- Google python style guide



## Key Features [continued]

- Save model for later use
- Single file configuration
- Pickle dump important objects
- No hard coding, can be used for n-number of features and any dataset



# Preprocessing Pipeline

- Removing duplicate rows for training
- Dropping columns with low standard deviation
- Dropping features that do not contribute to the learning
  - Found using correlation matrix
  - Brainstorming
  - Online research
- Handle outliers
  - Replace with median
- Deal with missing values
  - For categorical features, replace with mode
  - For numerical features, replace with median



# Preprocessing Pipeline [continued]

- Scale numeric features
  - Use min max normalization
- Encode categorical features
  - Label encoding
  - One-hot encoding
  - Feature hashing



# Model Building

- Logistic Regression
- L2 regularization
- Test size - 0.2
- Stratified shuffle split



# Planned Future Improvements

- Reiterate preprocessing pipeline
  - Encode categorical features
  - Missing values handling
    - Interpolation
- Feature engineering
  - Combining features
  - Defining our own features
- Ensembling of models
- More complex algorithms



Thank you!