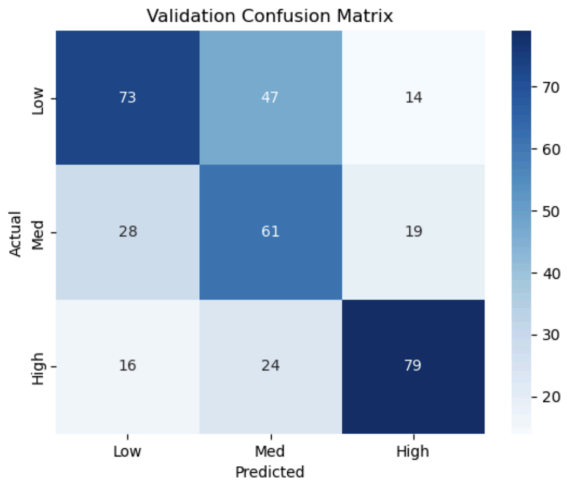
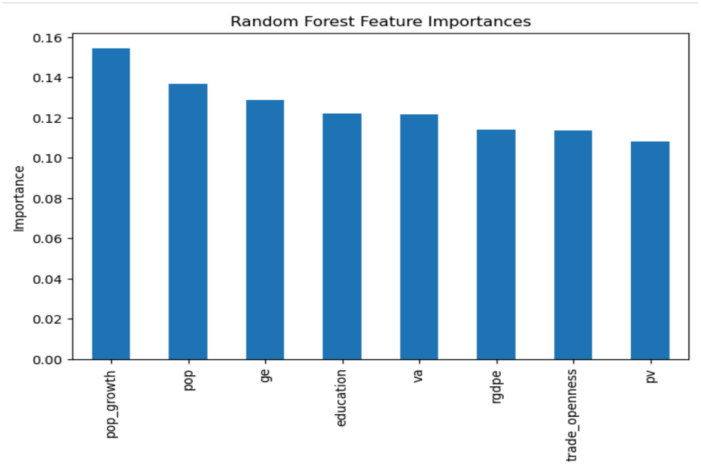
Working with Data:

- Set gdp_pc_growth as the target variable
  - Binned this variable into low, medium and high
  - Set purpose for model to classify countries into these bins
- Started with the following X variables
  - "va", "pv", "ge", "rq", "rl", "cc", "pop_growth", "education", "trade_openness", "inflation", "FDI", "investment_ratio", "rgdpe", "rgdpo", "pop"
- Created a correlation matrix to see which variables are heavily correlated
  - Removed variables with corr over .85
  - Ended with the following variables
    - `'va', 'pv', 'ge', 'pop_growth', 'education', 'trade_openness', 'rgdpe','pop'`

Model tuning and results:

- Used random forest in grid search to tune the hyperparameters
- Got the following model as the best
  - `Best params: {'max_depth': 15, 'min_samples_leaf': 4, 'min_samples_split': 2, 'n_estimators': 300}`
  - `Best CV score: 0.5306850714658209`
- Classification report on validation data
  - `Classification Report:`
  - `precision recall f1-score support`
  - `0 0.62 0.54 0.58 134`
  - `1 0.46 0.56 0.51 108`
  - `2 0.71 0.66 0.68 119`
  - `accuracy 0.59 361`
  - `macro avg 0.60 0.59 0.59 361`
  - `weighted avg 0.60 0.59 0.59 361`
-





-

Video of model usage with map of countries and shifting years

video of classification map

https://docs.google.com/presentation/d/1cl2ly1lm-jDVLOgbjtdJGocoVavdB0G3QxVo3bWEn8w/edit?usp=sharing