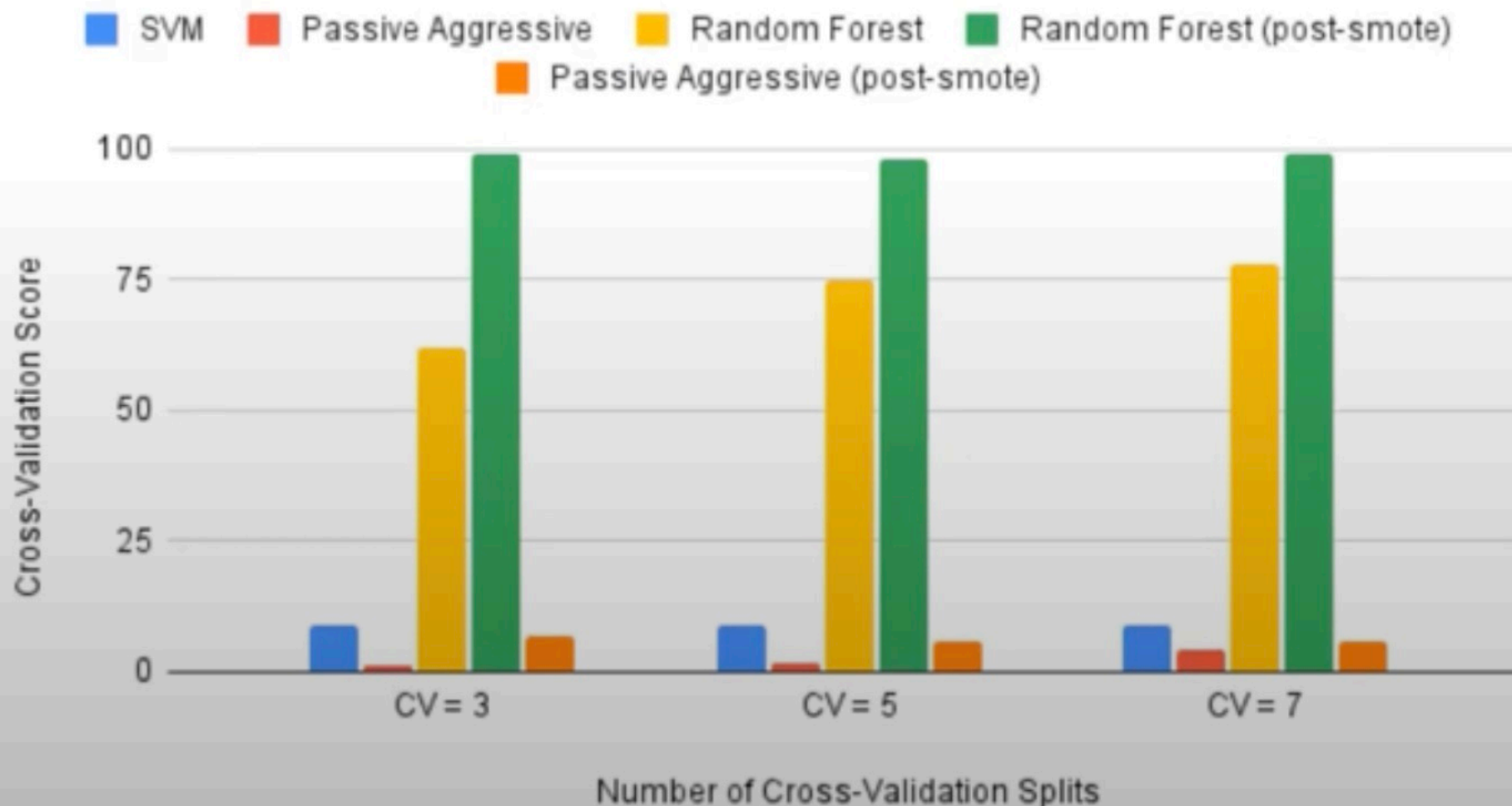# Experiments - Apply SMOTE

- Initial small imbalanced dataset would result in a "no tab" prediction most of the time

- Utilize SMOTE (Synthetic Minority Oversampling Technique) to balance the datasets by increasing minority class examples

- Gather larger data-set by running the extension over time also helped the initial imbalances

Post-SMOTE Cross Validation Scores

Legend: SVM, Passive Aggressive, Random Forest, Random Forest (post-smote), Passive Aggressive (post-smote)

Y-axis: Cross-Validation Score (0, 25, 50, 75, 100)

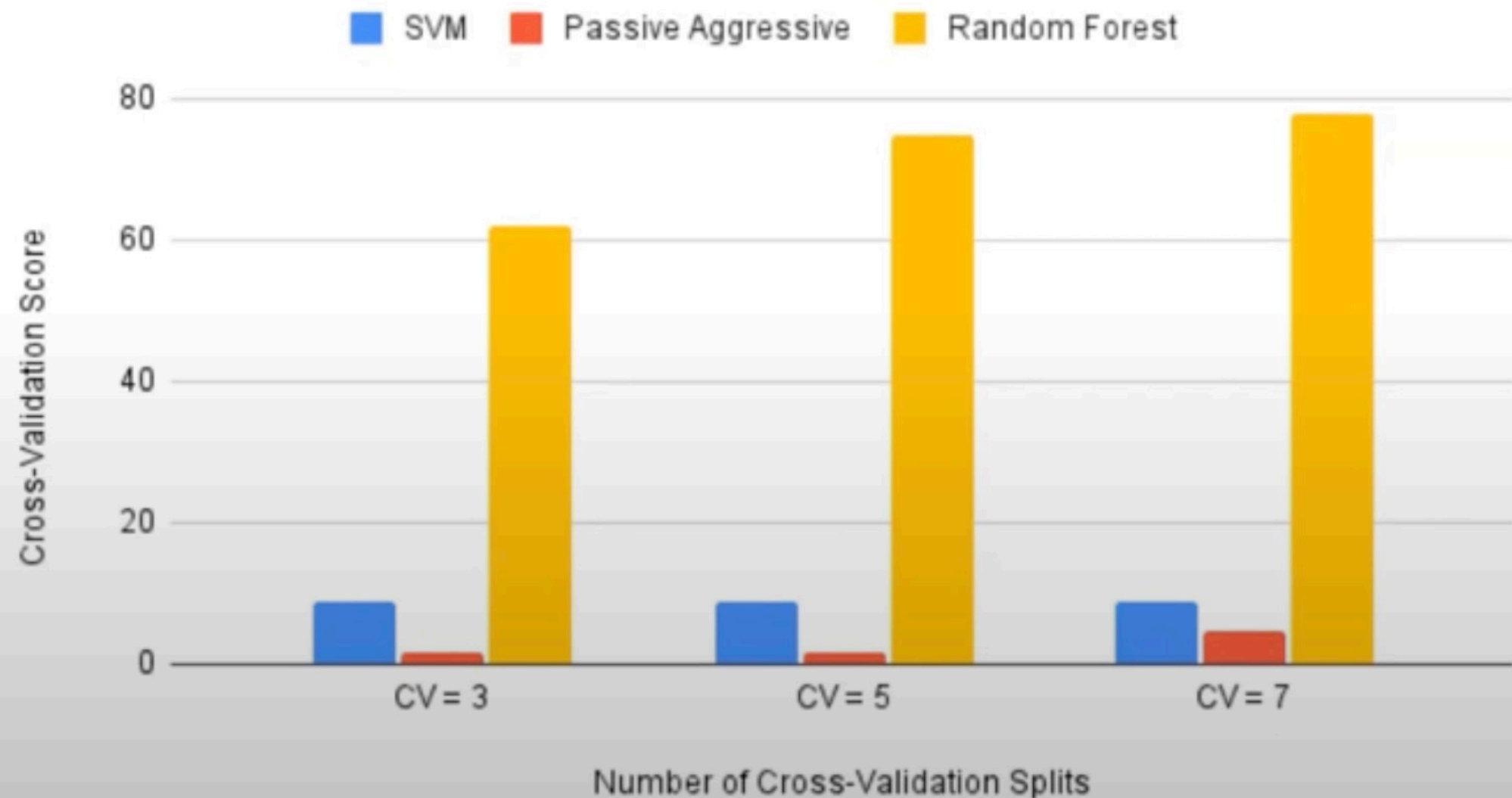X-axis: Number of Cross-Validation Splits (CV = 3, CV = 5, CV = 7)

Experiments - Time Interval for Prediction

- Find best time interval for the prediction
  - Test various intervals of 30 seconds, 1, 2, 5, and 10 minutes.
  - Pick 2-minute interval as it is relatively realistic
- Months and days are used for checking consistency between predicted URLs
- Milliseconds are used to ensure that two URLs are not opened within the same seconds

Experiments - ML Algorithms

- Explore Support Vector, Passive Aggressive, and Random Forest

- Use Sklearn K-fold cross-validation

Cross Validation Average Testing Scores

Legend: ■ SVM ■ Passive Aggressive ■ Random Forest

Y-axis: Cross-Validation Score (0, 20, 40, 60, 80)

X-axis: CV = 3, CV = 5, CV = 7

Number of Cross-Validation Splits

Experiments - Feature Importance

- Calculate SVM Feature Importance of Month, Day, Hour, Minute, and Second collected for Chrome Tab info