All precision, recall, and f1 scores are weighted averages. K-fold and stratified cross-validation scores are the average of 10 split results.  
  
The feature extraction part needed to be modified to add users to extracted features for a 30% user split to keep it for the test. The user has not retained during feature extraction previously. Feature extraction for the bonus part was added at the end of the ipynb file.

**Window 100 & 25% Overlap**

| Model | Accuracy | Precision | Recall | F1 Score | k-fold score | stratified score |
| --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.30 | 0.30 | 0.30 | 0.28 | 0.25 | 0.26 |
| Decision Trees | 0.51 | 0.51 | 0.51 | 0.51 | 0.601 | 0.603 |
| Random Forest | 0.64 | 0.64 | 0.64 | 0.64 | 0.72 | 0.72 |
| Gaussian Naïve Bayes | 0.22 | 0.23 | 0.22 | 0.19 | 0.13 | 0.13 |
| Support Vector Classifier | 0.56 | 0.55 | 0.56 | 0.55 | 0.275 | 0.274 |
| KNN | 0.52 | 0.54 | 0.52 | 0.51 | 0.360 | 0.365 |
| AdaBoost | 0.32 | 0.32 | 0.32 | 0.26 | 0.324 | 0.323 |
| XGBoost | 0.64 | 0.64 | 0.64 | 0.64 | 0.719 | 0.715 |
| Gradient Boost | 0.60 | 0.61 | 0.60 | 0.60 | 0.650 | 0.648 |
| Artificial Neural Networks (ANN) | 0.57 | 0.57 | 0.57 | 0.57 | 0.464 | 0.487 |

**Window 100 & 50% Overlap**

| Model | Accuracy | Precision | Recall | F1 Score | k-fold score | stratified score |
| --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.29 | 0.29 | 0.29 | 0.26 | 0.25 | 0.24 |
| Decision Trees | 0.53 | 0.54 | 0.53 | 0.54 | 0.62 | 0.63 |
| Random Forest | 0.68 | 0.69 | 0.68 | 0.68 | 0.76 | 0.77 |
| Gaussian Naïve Bayes | 0.26 | 0.26 | 0.26 | 0.22 | 0.13 | 0.13 |
| Support Vector Classifier | 0.58 | 0.58 | 0.58 | 0.57 | 0.29 | 0.29 |
| KNN | 0.55 | 0.56 | 0.55 | 0.53 | 0.37 | 0.37 |
| AdaBoost | 0.35 | 0.41 | 0.35 | 0.30 | 0.34 | 0.33 |
| XGBoost | 0.67 | 0.63 | 0.63 | 0.62 | 0.66 | 0.66 |
| Gradient Boost | 0.63 | 0.68 | 0.67 | 0.67 | 0.73 | 0.73 |
| Artificial Neural Networks (ANN) | 0.6 | 0.59 | 0.60 | 0.59 | 0.52 | 0.53 |

**Window 200 & 25% Overlap**

| Model | Accuracy | Precision | Recall | F1 Score | k-fold score | stratified score |
| --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.31 | 0.31 | 0.31 | 0.29 | 0.174 | 0.174 |
| Decision Trees | 0.63 | 0.63 | 0.63 | 0.62 | 0.72 | 0.72 |
| Random Forest | 0.47 | 0.48 | 0.47 | 0.47 | 0.591 | 0.593 |
| Gaussian Naïve Bayes | 0.25 | 0.26 | 0.25 | 0.22 | 0.26 | 0.259 |
| Support Vector Classifier | 0.55 | 0.55 | 0.55 | 0.55 | 0.25 | 0.255 |
| KNN | 0.49 | 0.53 | 0.49 | 0.48 | 0.349 | 0.349 |
| AdaBoost | 0.34 | 0.4 | 0.34 | 0.29 | 0.325 | 0.328 |
| XGBoost | 0.64 | 0.63 | 0.64 | 0.63 | 0.72 | 0.72 |
| Gradient Boost | 0.61 | 0.6 | 0.61 | 0.6 | 0.662 | 0.663 |
| Artificial Neural Networks (ANN) | 0.55 | 0.55 | 0.55 | 0.55 | 0.396 | 0.406 |

**Window 200 & 50% Overlap**

| Model | Accuracy | Precision | Recall | F1 Score | k-fold score | stratified score |
| --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.32 | 0.33 | 0.32 | 0.3 | 0.179 | 0.18 |
| Decision Trees | 0.54 | 0.53 | 0.54 | 0.53 | 0.631 | 0.633 |
| Random Forest | 0.68 | 0.68 | 0.68 | 0.67 | 0.766 | 0.768 |
| Gaussian Naïve Bayes | 0.24 | 0.27 | 0.24 | 0.21 | 0.153 | 0.152 |
| Support Vector Classifier | 0.59 | 0.59 | 0.59 | 0.58 | 0.277 | 0.277 |
| KNN | 0.53 | 0.55 | 0.53 | 0.51 | 0.369 | 0.371 |
| AdaBoost | 0.37 | 0.4 | 0.37 | 0.32 | 0.351 | 0.353 |
| XGBoost | 0.67 | 0.67 | 0.67 | 0.66 | 0.747 | 0.747 |
| Gradient Boost | 0.63 | 0.63 | 0.63 | 0.62 | 0.685 | 0.681 |
| Artificial Neural Networks (ANN) | 0.6 | 0.59 | 0.6 | 0.6 | 0.445 | 0.44 |

**Window 300 & 25% Overlap**

| Model | Accuracy | Precision | Recall | F1 Score | k-fold score | stratified score |
| --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.32 | 0.32 | 0.32 | 0.3 | 0.139 | 0.137 |
| Decision Trees | 0.47 | 0.46 | 0.47 | 0.46 | 0.585 | 0.58 |
| Random Forest | 0.62 | 0.62 | 0.62 | 0.61 | 0.724 | 0.726 |
| Gaussian Naïve Bayes | 0.3 | 0.31 | 0.3 | 0.27 | 0.149 | 0.151 |
| Support Vector Classifier | 0.54 | 0.54 | 0.54 | 0.53 | 0.236 | 0.236 |
| KNN | 0.48 | 0.52 | 0.48 | 0.46 | 0.345 | 0.344 |
| AdaBoost | 0.39 | 0.4 | 0.39 | 0.36 | 0.325 | 0.328 |
| XGBoost | 0.63 | 0.63 | 0.63 | 0.62 | 0.726 | 0.724 |
| Gradient Boost | 0.58 | 0.58 | 0.58 | 0.57 | 0.67 | 0.671 |
| Artificial Neural Networks (ANN) | 0.56 | 0.55 | 0.56 | 0.55 | 0.37 | 0.376 |

**Window 300 & 50% Overlap**

| Model | Accuracy | Precision | Recall | F1 Score | k-fold score | stratified score |
| --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.33 | 0.34 | 0.33 | 0.31 | 0.318 | 0.316 |
| Decision Trees | 0.5 | 0.5 | 0.5 | 0.5 | 0.635 | 0.637 |
| Random Forest | 0.66 | 0.66 | 0.66 | 0.66 | 0.771 | 0.77 |
| Gaussian Naïve Bayes | 0.28 | 0.31 | 0.28 | 0.25 | 0.275 | 0.275 |
| Support Vector Classifier | 0.58 | 0.57 | 0.58 | 0.57 | 0.256 | 0.255 |
| KNN | 0.51 | 0.54 | 0.51 | 0.49 | 0.372 | 0.372 |
| AdaBoost | 0.35 | 0.4 | 0.35 | 0.34 | 0.344 | 0.334 |
| XGBoost | 0.66 | 0.66 | 0.66 | 0.66 | 0.758 | 0.761 |
| Gradient Boost | 0.62 | 0.61 | 0.62 | 0.61 | 0.697 | 0.693 |
| Artificial Neural Networks (ANN) | 0.6 | 0.59 | 0.6 | 0.59 | 0.422 | 0.423 |

**Window 400 & 25% Overlap**

| Model | Accuracy | Precision | Recall | F1 Score | k-fold score | stratified score |
| --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.35 | 0.36 | 0.35 | 0.33 | 0.334 | 0.332 |
| Decision Trees | 0.46 | 0.46 | 0.46 | 0.46 | 0.582 | 0.58 |
| Random Forest | 0.61 | 0.61 | 0.61 | 0.6 | 0.725 | 0.723 |
| Gaussian Naïve Bayes | 0.3 | 0.33 | 0.3 | 0.28 | 0.289 | 0.289 |
| Support Vector Classifier | 0.54 | 0.54 | 0.54 | 0.53 | 0.222 | 0.223 |
| KNN | 0.47 | 0.51 | 0.47 | 0.45 | 0.349 | 0.351 |
| AdaBoost | 0.38 | 0.42 | 0.38 | 0.38 | 0.344 | 0.361 |
| XGBoost | 0.62 | 0.61 | 0.62 | 0.61 | 0.738 | 0.735 |
| Gradient Boost | 0.58 | 0.58 | 0.58 | 0.58 | 0.676 | 0.672 |
| Artificial Neural Networks (ANN) | 0.56 | 0.56 | 0.56 | 0.56 | 0.35 | 0.345 |

**Window 400 Overlap 50**

| Model | Accuracy | Precision | Recall | F1 Score | k-fold score | stratified score |
| --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.35 | 0.36 | 0.35 | 0.33 | 0.338 | 0.335 |
| Decision Trees | 0.51 | 0.51 | 0.51 | 0.51 | 0.628 | 0.63 |
| Random Forest | 0.66 | 0.65 | 0.66 | 0.65 | 0.776 | 0.775 |
| Gaussian Naïve Bayes | 0.31 | 0.34 | 0.31 | 0.28 | 0.291 | 0.29 |
| Support Vector Classifier | 0.59 | 0.59 | 0.59 | 0.59 | 0.253 | 0.252 |
| KNN | 0.49 | 0.53 | 0.49 | 0.48 | 0.376 | 0.379 |
| AdaBoost | 0.38 | 0.39 | 0.38 | 0.37 | 0.348 | 0.343 |
| XGBoost | 0.67 | 0.66 | 0.67 | 0.66 | 0.771 | 0.767 |
| Gradient Boost | 0.63 | 0.62 | 0.63 | 0.62 | 0.709 | 0.709 |
| Artificial Neural Networks (ANN) | 0.59 | 0.59 | 0.59 | 0.59 | 0.412 | 0.393 |

**Window 500 Overlap 25**

| Model | Accuracy | Precision | Recall | F1 Score | k-fold score | stratified score |
| --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.36 | 0.37 | 0.36 | 0.35 | 0.202 | 0.207 |
| Decision Trees | 0.46 | 0.46 | 0.46 | 0.45 | 0.59 | 0.589 |
| Random Forest | 0.6 | 0.6 | 0.6 | 0.6 | 0.732 | 0.733 |
| Gaussian Naïve Bayes | 0.29 | 0.32 | 0.29 | 0.27 | 0.172 | 0.173 |
| Support Vector Classifier | 0.55 | 0.55 | 0.55 | 0.54 | 0.218 | 0.214 |
| KNN | 0.45 | 0.5 | 0.45 | 0.43 | 0.353 | 0.356 |
| AdaBoost | 0.35 | 0.37 | 0.35 | 0.3 | 0.32 | 0.317 |
| XGBoost | 0.61 | 0.6 | 0.61 | 0.6 | 0.74 | 0.745 |
| Gradient Boost | 0.57 | 0.57 | 0.57 | 0.57 | 0.687 | 0.688 |
| Artificial Neural Networks (ANN) | 0.57 | 0.57 | 0.57 | 0.57 | 0.341 | 0.348 |

**Window 500 Overlap 50**

| Model | Accuracy | Precision | Recall | F1 Score | k-fold score | stratified score |
| --- | --- | --- | --- | --- | --- | --- |
| Logistic Regression | 0.37 | 0.37 | 0.37 | 0.36 | 0.203 | 0.206 |
| Decision Trees | 0.51 | 0.51 | 0.51 | 0.51 | 0.63 | 0.63 |
| Random Forest | 0.66 | 0.66 | 0.66 | 0.66 | 0.78 | 0.78 |
| Gaussian Naïve Bayes | 0.31 | 0.34 | 0.31 | 0.29 | 0.31 | 0.308 |
| KNN | 0.49 | 0.53 | 0.49 | 0.48 | 0.373 | 0.378 |
| Support Vector Classifier | 0.58 | 0.58 | 0.58 | 0.57 | 0.241 | 0.24 |
| AdaBoost | 0.37 | 0.4 | 0.37 | 0.36 | 0.376 | 0.376 |
| XGBoost | 0.67 | 0.67 | 0.67 | 0.67 | 0.779 | 0.78 |
| Gradient Boost | 0.62 | 0.62 | 0.62 | 0.62 | 0.718 | 0.716 |
| Artificial Neural Networks (ANN) | 0.56 | 0.56 | 0.56 | 0.56 | 0.368 | 0.343 |

**Bonus Section**

**Feature Extraction using TSFresh**

TSFresh was used to extract features for Windows 400 and 500, overlapping 25% and 50%. I have used ComprehensiveFCParameters(), which calculates the maximum number of features that TSFresh provides. Then imputed values on null cells in some cells might become zero in these processes if all values were null but this feature will be discarded during PCA.

Calculation results:

**Window 400 and 25% Overlap:** 2351 features with 19146 rows are generated

**Window 400 and 50% Overlap:** 2351 features with 28654 rows are generated

**Window 500 and 25% Overlap:** 2351 features with 15287 rows are generated

**Window 500 and 50% Overlap:** 2351 features with 22789 rows are generated

**PCA**

PCA was applied to calculate the number of components that retained a 95% variance. Among all of the extracted feature files, 5 features retained a 95% variance.

I applied the following models for window 500 with a 25% overlap

**LSTM:** accuracy 53.25%

**CNN:** accuracy 65.78%

**RNN:** accuracy 49.43%