Program Result

# Model Selection

## Without LDA

**Perceptron**

precision recall f1-score support

bullet 0.58 0.70 0.63 70728

irregular 0.80 0.70 0.75 257158

monthly 0.77 0.80 0.79 342717

weekly 0.07 0.28 0.12 602

micro avg 0.75 0.75 0.75 671205

macro avg 0.56 0.62 0.57 671205

weighted avg 0.76 0.75 0.76 671205

r2\_score mean\_squared\_error explained\_variance\_score support

bullet 0.089 0.086 0.095 70728

monthly 0.115 0.221 0.117 342717

weekly -3.26 0.004 -3.253 602

irregular 0.236 0.181 0.245 257158

micro avg -0.705 0.123 -0.699 671205

macro avg -0.326 0.246 0.0 671205

[[ 49632 5116 15965 15]

[ 9672 180943 64920 1623]

[ 26860 39832 275534 491]

[ 18 87 328 169]]

**Logistics Regression**

precision recall f1-score support

bullet 0.76 0.70 0.73 70728

irregular 0.85 0.78 0.81 257158

monthly 0.81 0.87 0.84 342717

weekly 0.00 0.00 0.00 602

micro avg 0.82 0.82 0.82 671205

macro avg 0.60 0.59 0.59 671205

weighted avg 0.82 0.82 0.82 671205

r2\_score mean\_squared\_error explained\_variance\_score support

bullet 0.412 0.055 0.413 70728

monthly 0.306 0.173 0.313 342717

weekly -0.001 0.001 0.0 602

irregular 0.418 0.138 0.422 257158

micro avg 0.284 0.092 0.287 671205

macro avg -0.225 0.184 0.0 671205

[[ 49476 3610 17642 0]

[ 3203 199911 54044 0]

[ 12731 31467 298519 0]

[ 0 70 532 0]]

**Decision Tree**

precision recall f1-score support

bullet 0.84 0.85 0.85 70728

irregular 0.89 0.91 0.90 257158

monthly 0.92 0.90 0.91 342717

weekly 0.74 0.71 0.72 602

micro avg 0.90 0.90 0.90 671205

macro avg 0.85 0.84 0.85 671205

weighted avg 0.90 0.90 0.90 671205

r2\_score mean\_squared\_error explained\_variance\_score support

bullet 0.652 0.033 0.652 70728

monthly 0.633 0.092 0.633 342717

weekly 0.456 0.0 0.456 602

irregular 0.685 0.074 0.686 257158

micro avg 0.606 0.05 0.607 671205

macro avg -0.111 0.1 0.0 671205

[[ 60007 2576 8132 13]

[ 2457 234935 19656 110]

[ 8833 24962 308893 29]

[ 3 146 26 427]]

**Random Forest**

precision recall f1-score support

bullet 0.85 0.87 0.86 70728

irregular 0.90 0.92 0.91 257158

monthly 0.92 0.91 0.91 342717

weekly 0.76 0.66 0.71 602

micro avg 0.91 0.91 0.91 671205

macro avg 0.86 0.84 0.85 671205

weighted avg 0.91 0.91 0.91 671205

r2\_score mean\_squared\_error explained\_variance\_score support

bullet 0.681 0.03 0.681 70728

monthly 0.653 0.087 0.653 342717

weekly 0.455 0.0 0.455 602

irregular 0.706 0.07 0.706 257158

micro avg 0.624 0.047 0.624 671205

macro avg -0.103 0.093 0.0 671205

[[ 61183 1880 7664 1]

[ 2348 235401 19298 111]

[ 8279 22888 311536 14]

[ 1 167 34 400]]

## **With LDA**

**Perceptron**

precision recall f1-score support

bullet 0.52 0.78 0.62 70728

irregular 0.86 0.73 0.79 257158

monthly 0.81 0.82 0.82 342717

weekly 0.00 0.00 0.00 602

micro avg 0.78 0.78 0.78 671205

macro avg 0.55 0.58 0.56 671205

weighted avg 0.80 0.78 0.79 671205

r2\_score mean\_squared\_error explained\_variance\_score support

bullet -0.051 0.099 -0.021 70728

monthly 0.252 0.187 0.253 342717

weekly -0.004 0.001 -0.003 602

irregular 0.368 0.149 0.383 257158

micro avg 0.141 0.109 0.153 671205

macro avg -0.279 0.218 0.0 671205

[[ 55260 2280 13187 1]

[ 18217 187016 51924 1]

[ 32305 27860 282552 0]

[ 487 0 115 0]]

**Logistics Regression**

precision recall f1-score support

bullet 0.73 0.63 0.67 70728

irregular 0.86 0.76 0.81 257158

monthly 0.78 0.88 0.83 342717

weekly 0.00 0.00 0.00 602

micro avg 0.81 0.81 0.81 671205

macro avg 0.59 0.57 0.58 671205

weighted avg 0.81 0.81 0.80 671205

r2\_score mean\_squared\_error explained\_variance\_score support

bullet 0.325 0.064 0.327 70728

monthly 0.254 0.186 0.268 342717

weekly -0.001 0.001 0.0 602

irregular 0.417 0.138 0.425 257158

micro avg 0.249 0.097 0.255 671205

macro avg -0.241 0.194 0.0 671205

[[ 44229 3597 22902 0]

[ 1592 196313 59253 0]

[ 14639 27877 300201 0]

[ 0 157 445 0]]

**Decision Tree**

precision recall f1-score support

bullet 0.78 0.79 0.79 70728

irregular 0.84 0.85 0.85 257158

monthly 0.86 0.85 0.86 342717

weekly 0.53 0.48 0.50 602

micro avg 0.85 0.85 0.85 671205

macro avg 0.75 0.74 0.75 671205

weighted avg 0.85 0.85 0.85 671205

r2\_score mean\_squared\_error explained\_variance\_score support

bullet 0.524 0.045 0.524 70728

monthly 0.422 0.144 0.422 342717

weekly 0.054 0.001 0.054 602

irregular 0.494 0.12 0.494 257158

micro avg 0.374 0.078 0.374 671205

macro avg -0.183 0.155 0.0 671205

[[ 56074 3271 11364 19]

[ 3310 219553 34113 182]

[ 12154 39216 291291 56]

[ 4 246 62 290]]

**Random Forest**

precision recall f1-score support

bullet 0.81 0.83 0.82 70728

irregular 0.86 0.87 0.87 257158

monthly 0.88 0.87 0.88 342717

weekly 0.64 0.45 0.52 602

micro avg 0.87 0.87 0.87 671205

macro avg 0.80 0.76 0.77 671205

weighted avg 0.87 0.87 0.87 671205

r2\_score mean\_squared\_error explained\_variance\_score support

bullet 0.597 0.038 0.597 70728

monthly 0.5 0.125 0.5 342717

weekly 0.194 0.001 0.194 602

irregular 0.566 0.103 0.566 257158

micro avg 0.464 0.067 0.464 671205

macro avg -0.154 0.133 0.0 671205

[[ 58929 2298 9493 8]

[ 2817 224485 29733 123]

[ 10888 33651 298158 20]

[ 3 277 54 268]]

**KNN**

precision recall f1-score support

bullet 0.81 0.86 0.83 70728

irregular 0.87 0.89 0.88 257158

monthly 0.90 0.87 0.89 342717

weekly 0.58 0.39 0.47 602

micro avg 0.88 0.88 0.88 671205

macro avg 0.79 0.75 0.77 671205

weighted avg 0.88 0.88 0.88 671205

r2\_score mean\_squared\_error explained\_variance\_score support

bullet 0.616 0.036 0.616 70728

monthly 0.538 0.115 0.539 342717

weekly 0.107 0.001 0.107 602

irregular 0.6 0.095 0.6 257158

micro avg 0.465 0.062 0.466 671205

macro avg -0.141 0.124 0.0 671205

[[ 60979 1892 7826 31]

[ 3064 228710 25264 120]

[ 11509 32802 298388 18]

[ 3 320 45 234]]

**SVM (first 10,000 for train and 1,000 for test)**

precision recall f1-score support

bullet 0.69 0.37 0.48 1019

irregular 0.83 0.82 0.83 3886

monthly 0.78 0.86 0.82 5081

weekly 0.00 0.00 0.00 14

micro avg 0.79 0.79 0.79 10000

macro avg 0.57 0.51 0.53 10000

weighted avg 0.79 0.79 0.78 10000

r2\_score mean\_squared\_error explained\_variance\_score support

bullet 0.107 0.082 0.132 1019

monthly 0.212 0.197 0.223 5081

weekly -0.001 0.001 0.0 14

irregular 0.433 0.135 0.434 3886

micro avg 0.188 0.104 0.197 10000

macro avg -0.262 0.207 0.0 10000

[[ 373 40 606 0]

[ 52 3201 633 0]

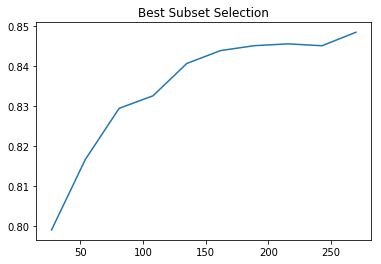
[ 119 609 4353 0]

[ 0 12 2 0]]

# **Advanced Training**

## Feature Extraction

**Best Subset**



## **Filtering**

**Filtering, Decision Tree to filter out weekly and then Decision Tree to identify bullet, monthly and irregular if not weekly.**

precision recall f1-score support

bullet 0.84 0.85 0.85 70728

irregular 0.89 0.91 0.90 257158

monthly 0.92 0.90 0.91 342717

weekly 0.71 0.71 0.71 602

micro avg 0.90 0.90 0.90 671205

macro avg 0.84 0.84 0.84 671205

weighted avg 0.90 0.90 0.90 671205

r2\_score mean\_squared\_error explained\_variance\_score support

bullet 0.653 0.033 0.653 70728

monthly 0.632 0.092 0.632 342717

weekly 0.418 0.001 0.418 602

irregular 0.685 0.074 0.685 257158

micro avg 0.597 0.05 0.597 671205

macro avg -0.111 0.1 0.0 671205

[[ 59960 2581 8177 10]

[ 2410 234906 19705 137]

[ 8788 25008 308893 28]

[ 7 142 26 427]]

**Without Filtering, Single Decision Tree**

precision recall f1-score support

bullet 0.84 0.85 0.85 70728

irregular 0.89 0.91 0.90 257158

monthly 0.92 0.90 0.91 342717

weekly 0.74 0.71 0.72 602

micro avg 0.90 0.90 0.90 671205

macro avg 0.85 0.84 0.85 671205

weighted avg 0.90 0.90 0.90 671205

r2\_score mean\_squared\_error explained\_variance\_score support

bullet 0.652 0.033 0.652 70728

monthly 0.633 0.092 0.633 342717

weekly 0.456 0.0 0.456 602

irregular 0.685 0.074 0.686 257158

micro avg 0.606 0.05 0.607 671205

macro avg -0.111 0.1 0.0 671205

[[ 60007 2576 8132 13]

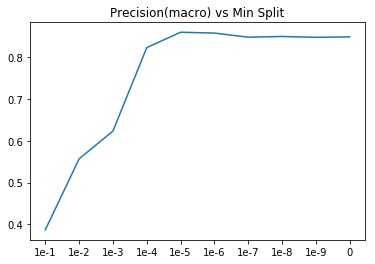
[ 2457 234935 19656 110]

[ 8833 24962 308893 29]

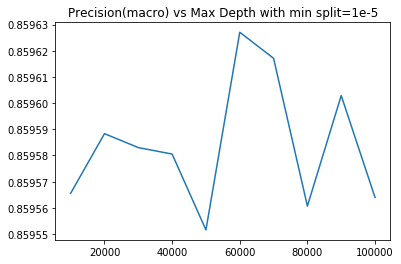
[ 3 146 26 427]]

## **Tune Parameters**

**Min Split Decrease**



**Max Depth**



# Final Result

## Performance

precision recall f1-score support

bullet 0.85 0.88 0.86 70728

irregular 0.92 0.92 0.92 257158

monthly 0.93 0.92 0.92 342717

weekly 0.74 0.84 0.79 602

micro avg 0.92 0.92 0.92 671205

macro avg 0.86 0.89 0.87 671205

weighted avg 0.92 0.92 0.92 671205

r2\_score mean\_squared\_error explained\_variance\_score support

bullet 0.693 0.029 0.693 70728

monthly 0.688 0.078 0.688 342717

weekly 0.546 0.0 0.546 602

irregular 0.738 0.062 0.738 257158

micro avg 0.666 0.042 0.666 671205

macro avg -0.092 0.085 0.0 671205

[[ 61995 1684 7034 15]

[ 2531 236837 17638 152]

[ 8159 19444 315103 11]

[ 5 65 25 507]]

## Tree

