## 4.3 Trip Purpose Classification Models

### 4.3.1 Preliminary results

To better account for the spatial and temporal irregularity and dependency (identified in 4.2), we explicitly divide the data into 5 unique subsets and these are detailed in **Table 4.8**.

***Table 4.8*** *Subsets used in the building of classification models*

|  |  |  |
| --- | --- | --- |
| *Subset* | *N* | *% of total trips* |
| All Data | 71,801 | 100 |
| Rush Hour | 40,945 | 56.1 |
| Off-Peak | 32,084 | 43.9 |
| City | 68,275 | 93.5 |
| Non-city | 4,754 | 6.5 |

The results from the hyperparameter tuning are as follows

Support vector model built using the 0.1

'alpha': 0.01,

'hidden\_layer\_sizes': (50, 50, 50),

'max\_iter': 500,

'solver': 'lbfgs'

For the purposes of the classification models, the data is normalised for the purpose of the

Cs = [0.1, 1, 10]

gammas = [0.01, 0.1, 1]

Best == {'C': 0.1, 'gamma': 0.01}

tested for alpha == 1e-05, 0.0001, 0.05

'sgd', 'adam', 'lbfgs'

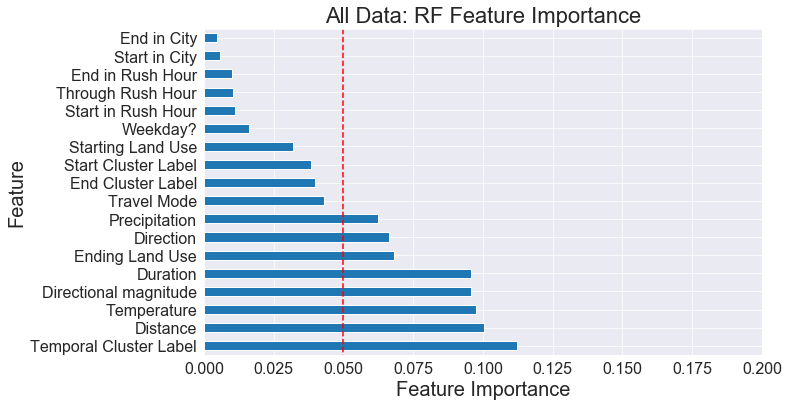
Alpha == learning rate

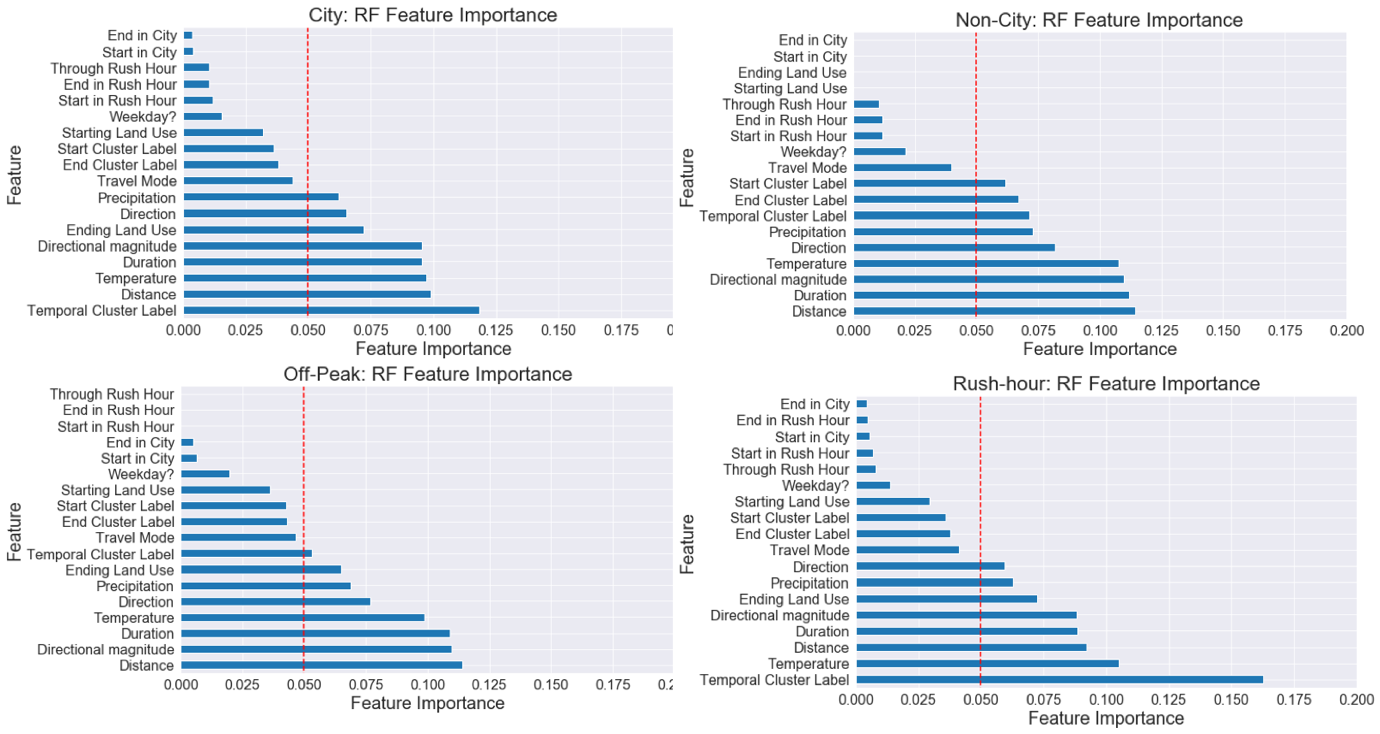
'alpha': 0.01,

'hidden\_layer\_sizes': (50, 50, 50),

'max\_iter': 500,

'solver': 'lbfgs'





***Figure 4.X*** *Feature importance from a Random Forest Regression model (the red line indicates model features which will be removed)*

### 4.3.2 Classifier results

We find that all the models are accurate 6,857 trips and all inaccurate in 10,317

Results:

***Table 4.X*** *Overall Accuracy in the models*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Classifier* | *5-fold cross-validation* | | *% of trips where prediction made* | *Accuracy* |
| *Accuracy (%)* | *Range (%)* |
| Random Forest | 0.59 | 0.9 | 56.4 | 0.67 |
| SVC | 0.48 | 1.2 | 100 | 0.48 |
| MLP | 0.43 | 1.5 | 60.8 | 0.69 |

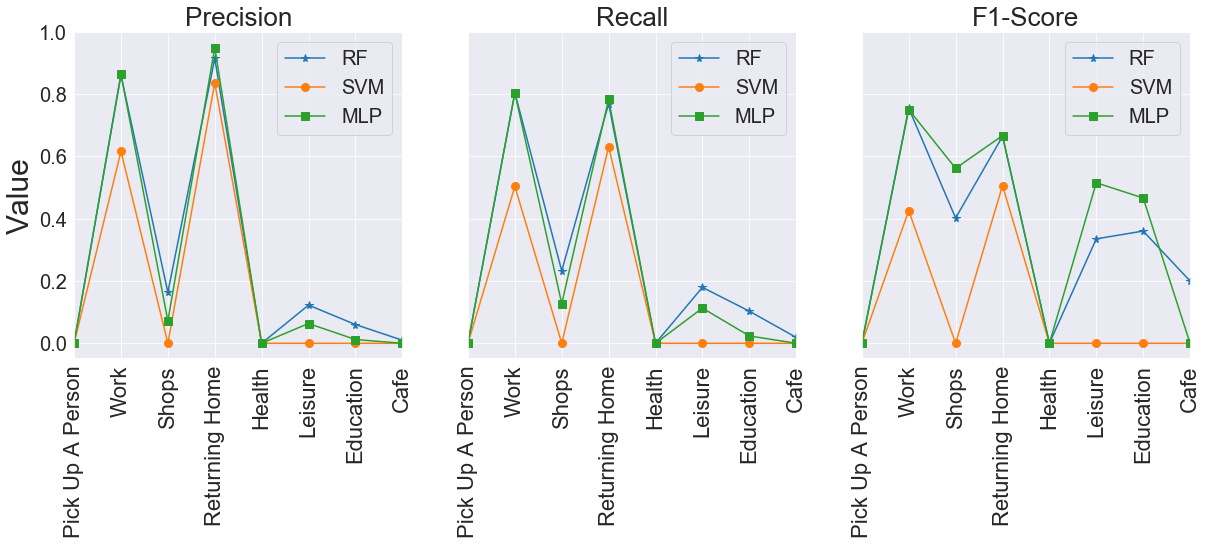
Recall -> correct for the class vs all for the class

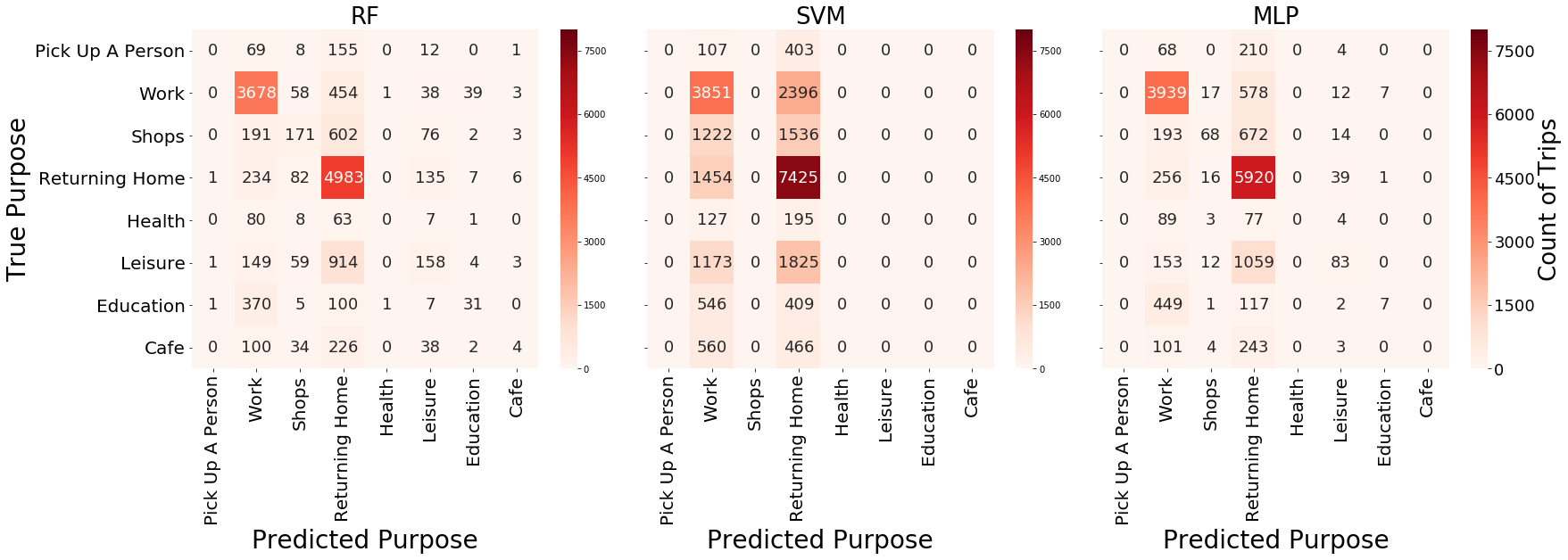
Precision -> correct for the class vs correct and incorrect for the class

NN precision above half correct (above 0.5) for 4 , meaning the model was slightly more confident when making prediction for these

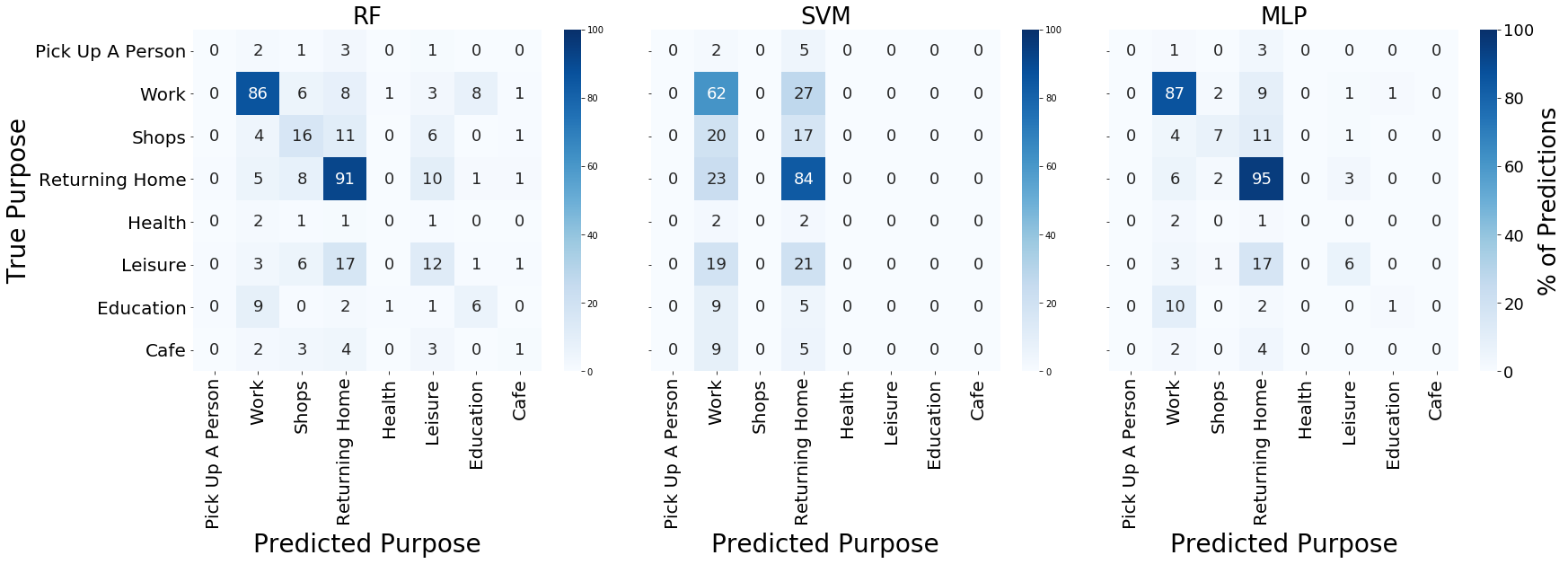
**Table 4.X** Results from the classification broken down by class of trip purpose (values above 0.5 are shown in **bold**).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Random Forest Classifier | | | | |
|  | *Precision* | *Recall* | *F1-score* | *Support* |
|  | 0 | 0 | 0 | 245 |
| *Work* | **0.76** | **0.86** | **0.80** | 4271 |
| *Shops* | 0.40 | 0.16 | 0.23 | 1045 |
| *Returning Home* | **0.66** | **0.91** | **0.77** | 5548 |
| *Health* | 0 | 0 | 0 | 159 |
| *Leisure* | 0.34 | 0.12 | 0.18 | 1288 |
| *Education* | 0.36 | 0.06 | 0.10 | 515 |
| *Cafe* | 0.20 | 0.01 | 0.02 | 404 |
| *Accuracy* |  |  | **0.67** | 13375 |
| Support Vector Classifier | | | | |
|  | Precision | Recall | F1-score | Support |
| *Pick up a person* | 0 | 0 | 0 | 510 |
| *Work* | 0.43 | **0.62** | **0.50** | 6247 |
| *Shops* | 0 | 0 | 0 | 2758 |
| *Returning Home* | **0.51** | **0.84** | **0.63** | 8879 |
| *Health* | 0 | 0 | 0 | 332 |
| *Leisure* | 0 | 0 | 0 | 2998 |
| *Education* | 0 | 0 | 0 | 955 |
| *Cafe* | 0 | 0 | 0 | 1026 |
| *Accuracy* |  |  | 0.48 | 23695 |
| Multilayer Perceptron | | | | |
|  | *Precision* | *Recall* | *F1-score* | *Support* |
| *Pick up a person* | 0 | 0 | 0 | 282 |
| *Work* | **0.75** | **0.87** | **0.80** | 4553 |
| *Shops* | **0.56** | 0.07 | 0.13 | 947 |
| *Returning Home* | **0.67** | **0.95** | **0.78** | 6232 |
| *Health* | 0 | 0 | 0 | 173 |
| *Leisure* | **0.52** | 0.06 | 0.11 | 1307 |
| *Education* | 0.47 | 0.01 | 0.02 | 576 |
| *Cafe* | 0 | 0 | 0 | 351 |
| *Accuracy* |  |  | **0.69** | 14421 |

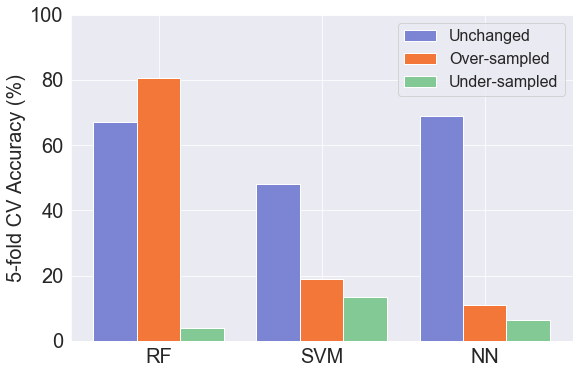




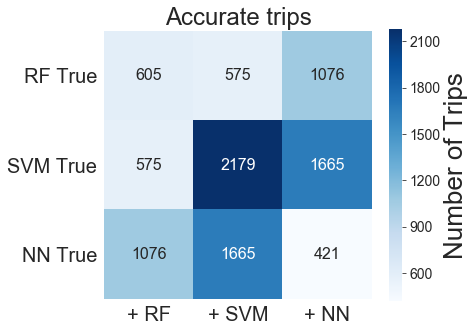
***Figure 4.X*** *True vs predicted purpose confusion matrices for each classifier*



***Figure 4.X*** *True vs predicted purpose confusion matrices for each classifier (as % of total predictions)*

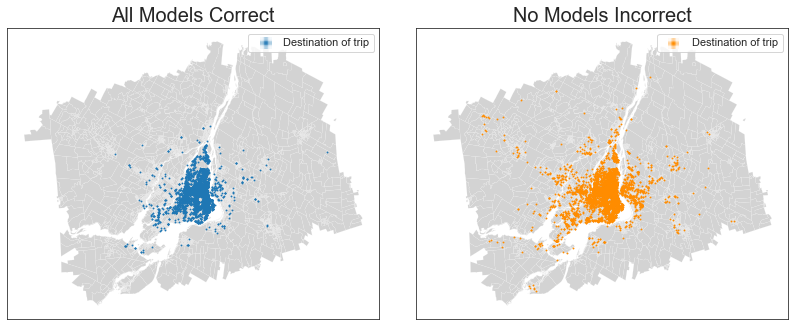


***Figure 4.X*** *Bar plots comparing the cross-validation accuracy*



***Figure 4.X*** *Matrix showing the amount of correctly identified trips in each one of the models compared to the others*

When comparing the spatial

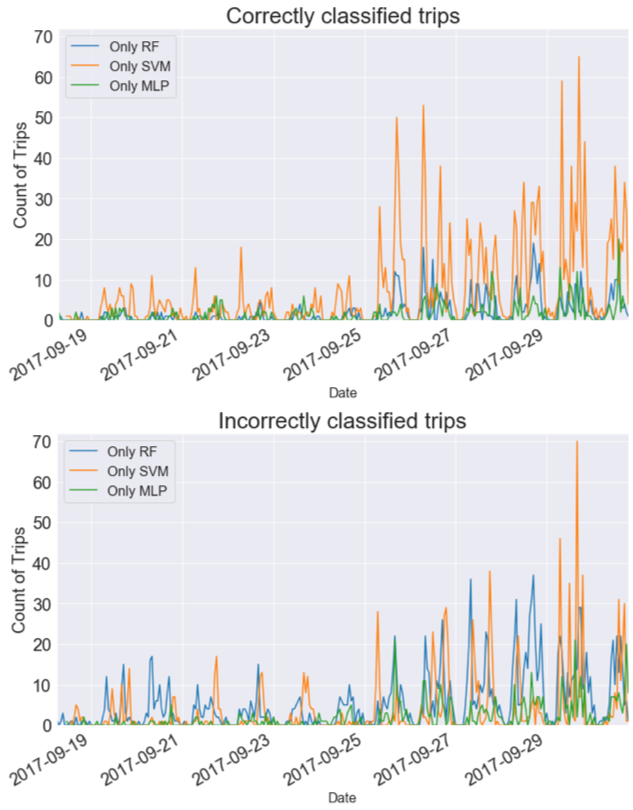


***Figure 4.X*** *Destination of trips that are classified correctly (left) and incorrectly (right) by all the classifiers.*

Temporality no true -3.083525711470756, 0.027796626120368234

All true -3.9825091284053635,

0.0015049193012149386,



***Figure 4.X*** *Amount of correctly (top) and incorrectly (bottom) classified trips by each of the classifiers across the study period.*