**Predicting Blood Glucose Levels of Diabetes Patients**

We create Machine Learning Algorithms in Python to predict the blood glucose levels of diabetes patients. The work is done with the UCI Diabetes data set consisting of 70 sets of data recorded on diabetes patients. We apply a three-stage evolution model as in the article "Smartphone-based personalized blood glucose prediction" written by Juan Li and Chandima Fernando:

(1) Time Series prediction model based on patient data: apply time series forecasting with Support Vector Machine (SVM), Decision Tree (DT) and Random Forest (RF).

(2) Pooled-panel regression model (PPD): apply linear regression algorithm to panel data in order to increase the sample size.

(3) Pre-clustered personalized regression model: to remove patients with different behaviour.

We basically follow the article except for that we do not make use of stage (3) pre-clustering. Instead, I think it is faster and easier to remove rare patients and behaviours manually.

The metrics used for testing are 'mean absolute error' (MAE) and 'root-mean-squared error' (RMSE). Furthermore, we use cross validation to check the accuracies of these models. The accuracies of these models can be summarized as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | SVM | DT | RF | PPD |
| MAE | 61.11 | 48.54 | 44.88 | 42.97 |
| RMSE | 68.64 | 56.07 | 52.70 | 50.48 |

In conclusion, PPD provides the best predictions. Among the three time series methods mentioned above, Random Forest gives the best predictions. Moreover, we develop more the RF and PPD models to get better predictions as below. Note that the new models are not recommended since it takes much more time to train them.

|  |  |  |
| --- | --- | --- |
|  | RF-develop | PPD-develop |
| MAE | 41.65 | 39.78 |
| RMSE | 41.65 | 39.78 |