**COMP9321 – Assignment3**

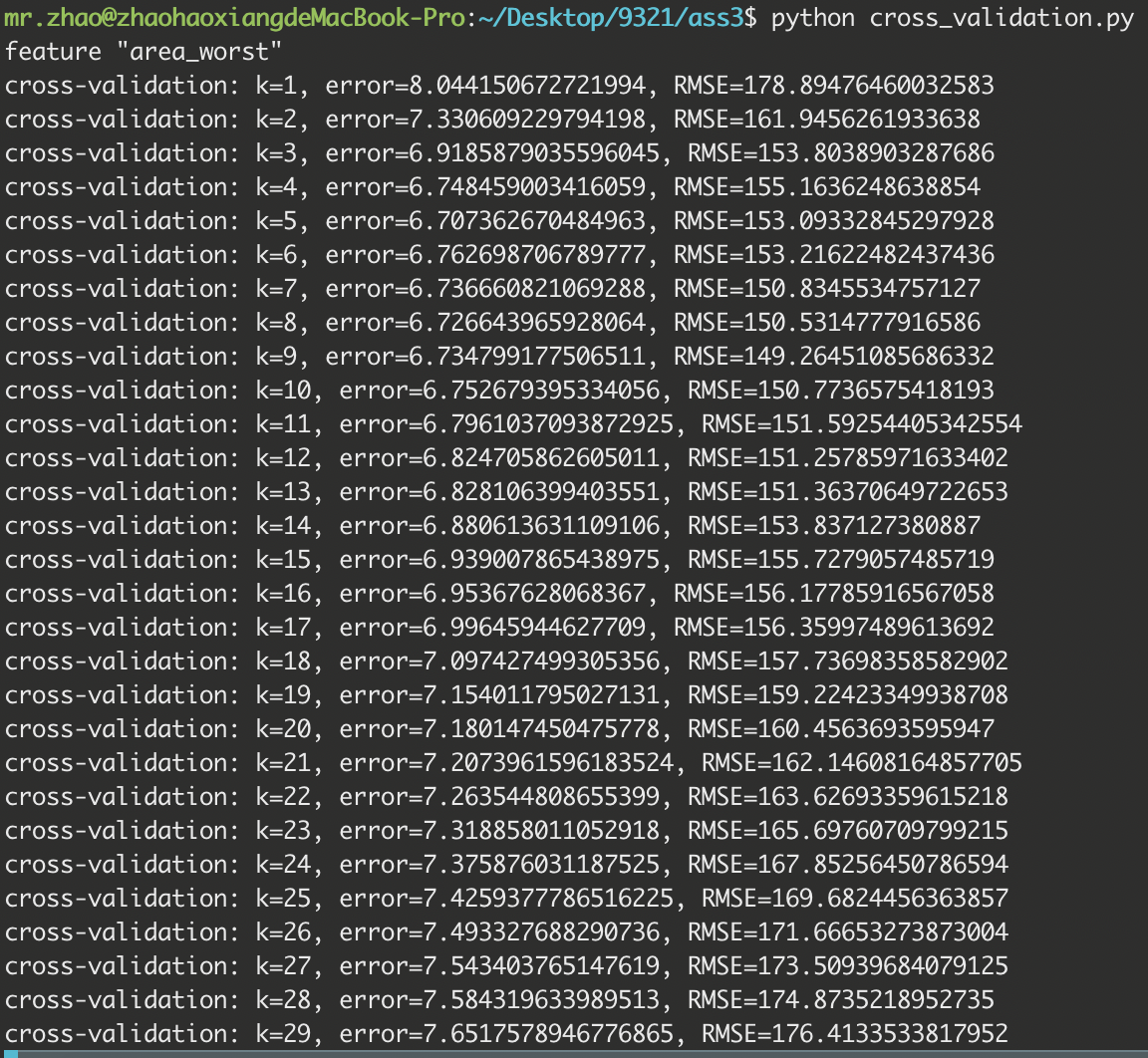
**Team FreshCoke**

**‐ Machine Learning:**

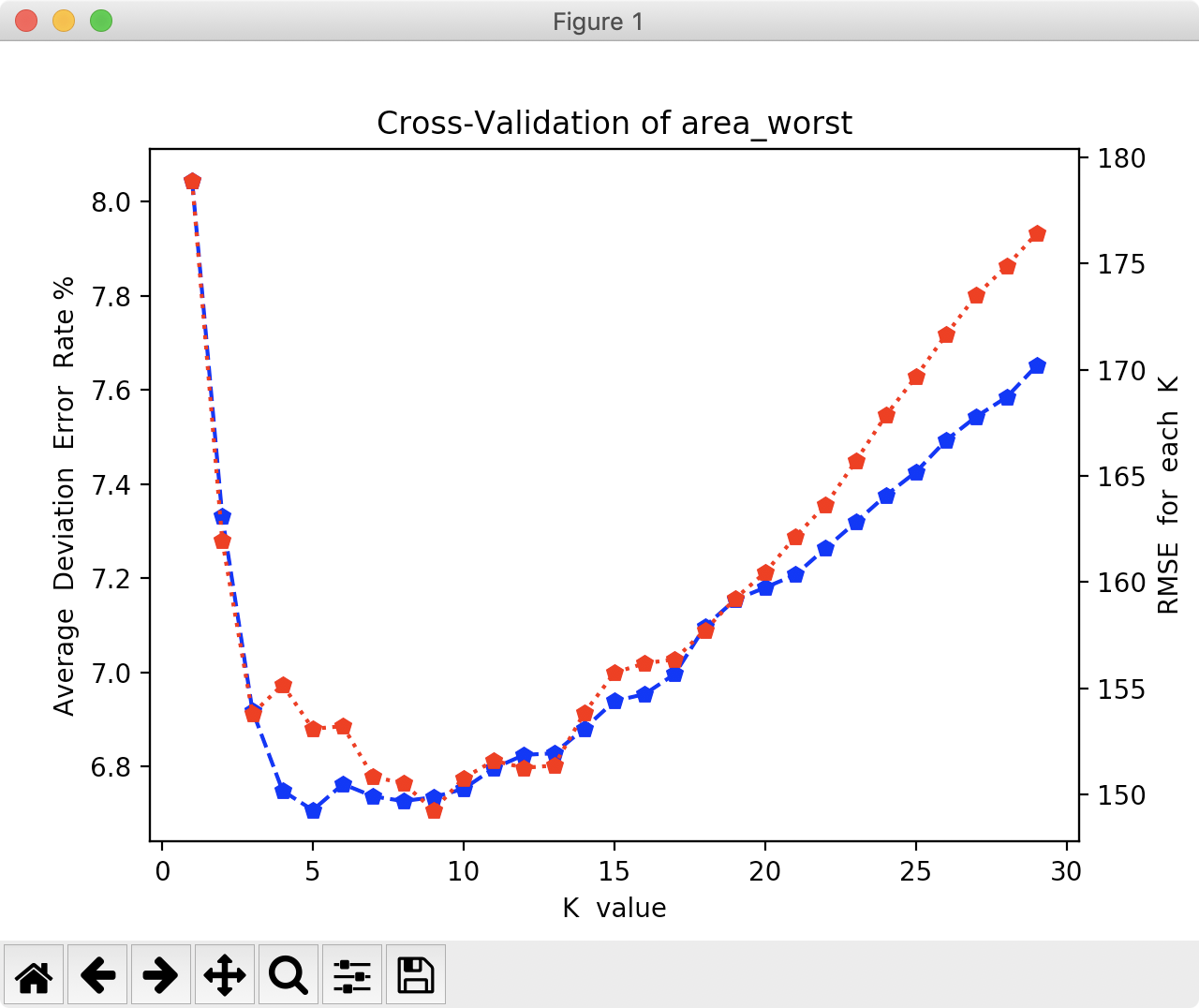
**In the classification part**, we test different ML models and finally choose SVM (support vector machine) as the model to train the dataset. The SVM cross-one-validation result can reach a high accuracyat **95.43%**.

We have also tried KNN to fit this dataset, but it shows lower accuracy than SVM. Random Forest can achieve a decent as well, but we are more familiar with SVM than this.

**In the prediction part,** we used K-NN as the algorithm to predict values of the features. Due to the nature of SVM, it is more suitable for classification but not the prediction. It is important to choose the best K number for the KNN to fit datasets, and that is why we use cross-validation again.



After calculating error and RMSE (Root Mean Square Error) of 30 different Ks, the best K for feature “Area-Mean” is 9 (lower error and RMSE). Same as “Area-Mean”, we can finally find the best K for our prediction system is 10.



**‐ Table of Tasks (until week 12):**

|  |  |
| --- | --- |
| Done | To do |
| Acquire dataset | Adjust webpage detail (more user friendly) |
| Data cleaning | UI polish |
| Machine Learning (classification and prediction) |  |
| RESTful API server |  |
| Back-end frame work |  |
| Web/UI design (structure and layout) |  |
| User Interaction Design |  |

**‐ Webpage review:**

