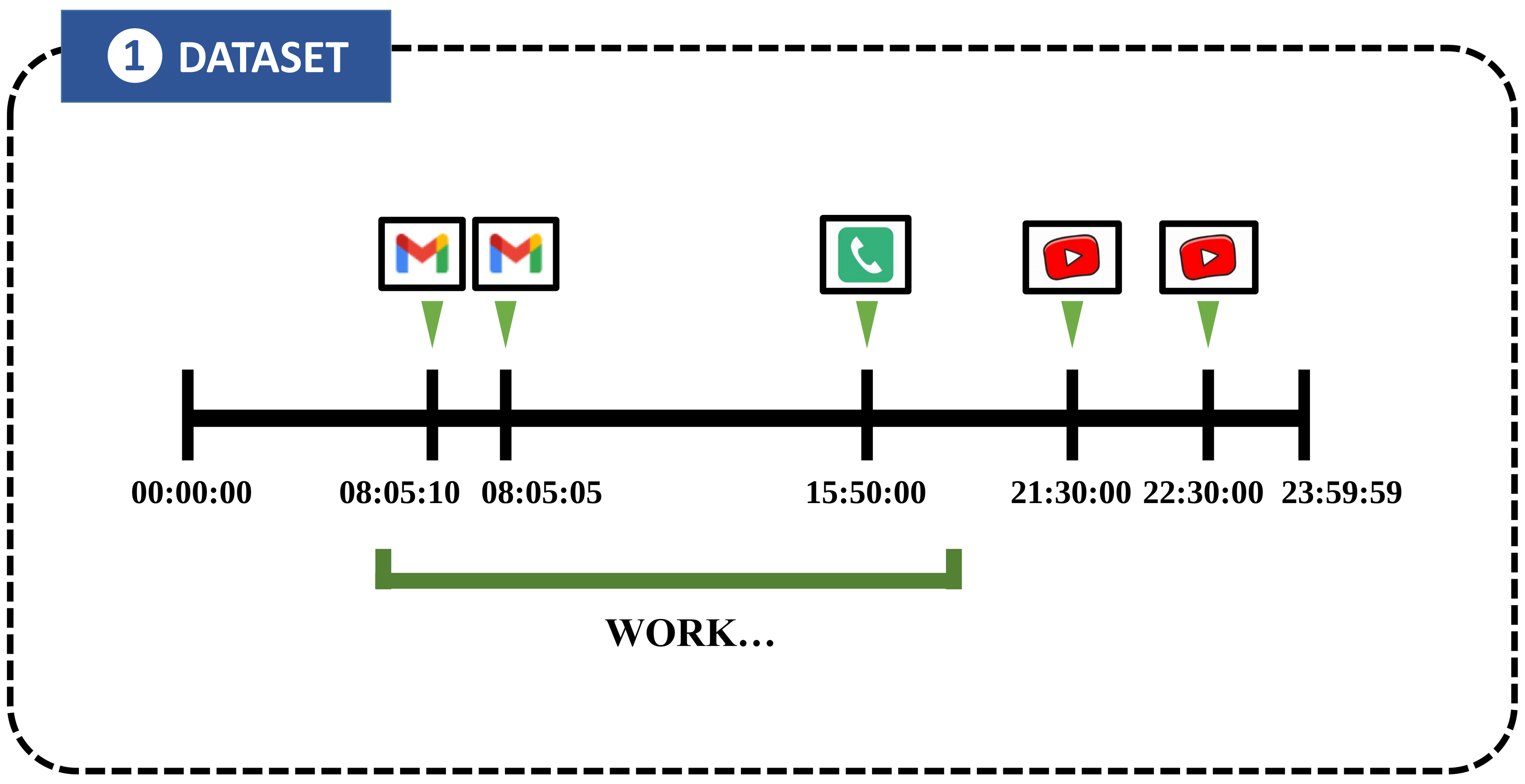
Individual working efficiency analysis by cellphone usage status

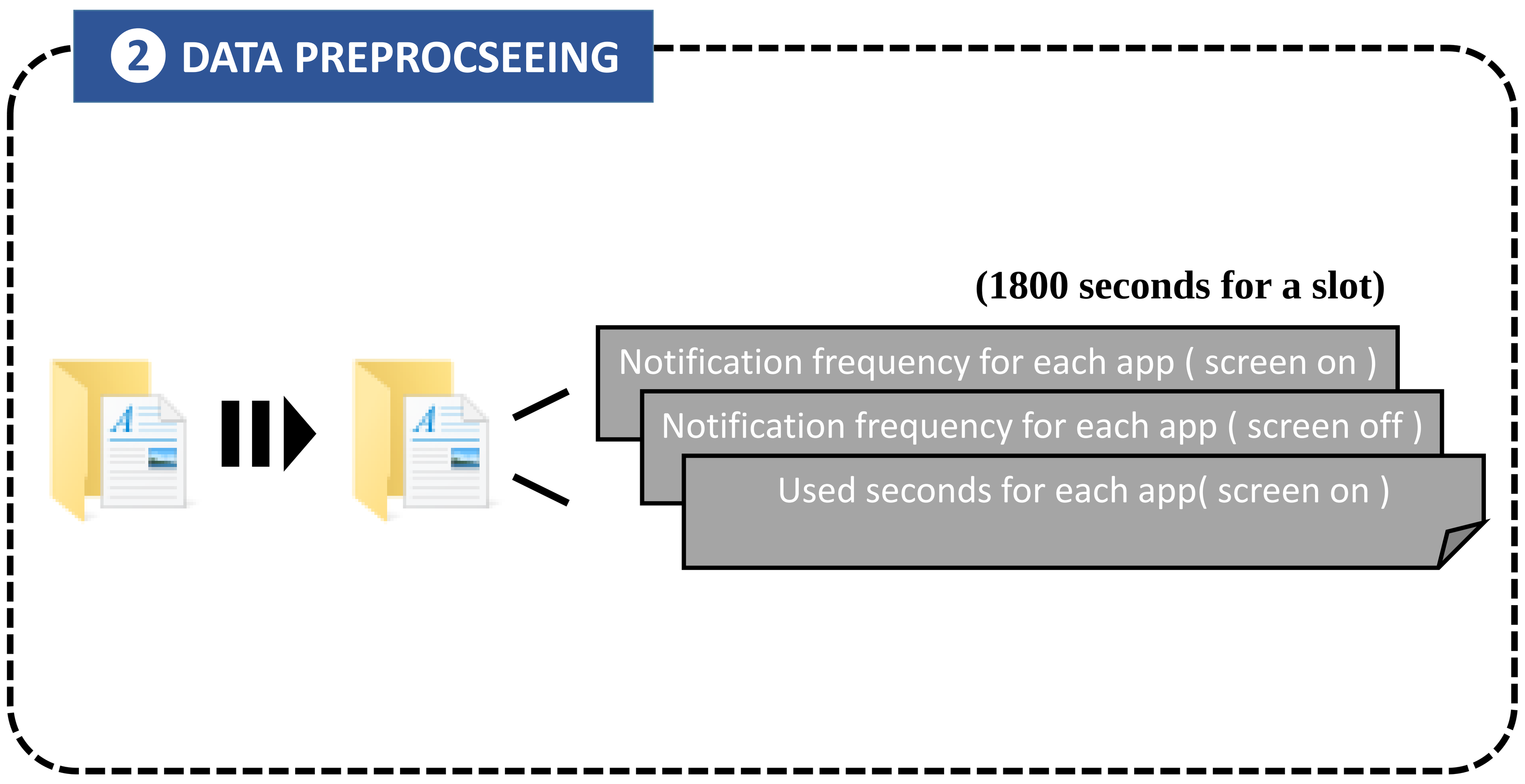
Wei-Shan Chen (陳葦珊), *Institute of Statistics, National Yang Ming Chiao Tung University, Taiwan* (sandy58976689@gmail.com)
Yu-Hsuan Lin (林煜軒), *Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan* (yuhsuanlin@nhri.edu.tw)
Hung-Hsun Chen (陳泓勳), *Center for Teaching and Learning Development, National Yang Ming Chiao Tung University, Taiwan* (hhchen@nctu.edu.tw)
Chih-Hsuan Chen (陳芷萱), *Project Assignment, National Yang Ming Chiao Tung University, Taiwan* (qopx2233@gmail.com)
Yu-Ting Cheng (鄭宇廷), *Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan* (rainsstop@gmail.com)
Ting-Yun Chiu (邱庭芸), *Institute of Population Health Sciences, National Health Research Institutes, Miaoli, Taiwan* (tyc090918@nhri.edu.tw)
Wei-Shiang Chen (陳偉翔), *Institute of Statistics, National Yang Ming Chiao Tung University* (wschen.st06g@nctu.edu.tw)
Henry Horng-Shing Lu (盧鴻興), *Institute of Statistics, National Yang Ming Chiao Tung University, Taiwan* (hslu@stat.nctu.edu.tw)
Advisor: Prof. Henry Horng-Shing Lu (盧鴻興)

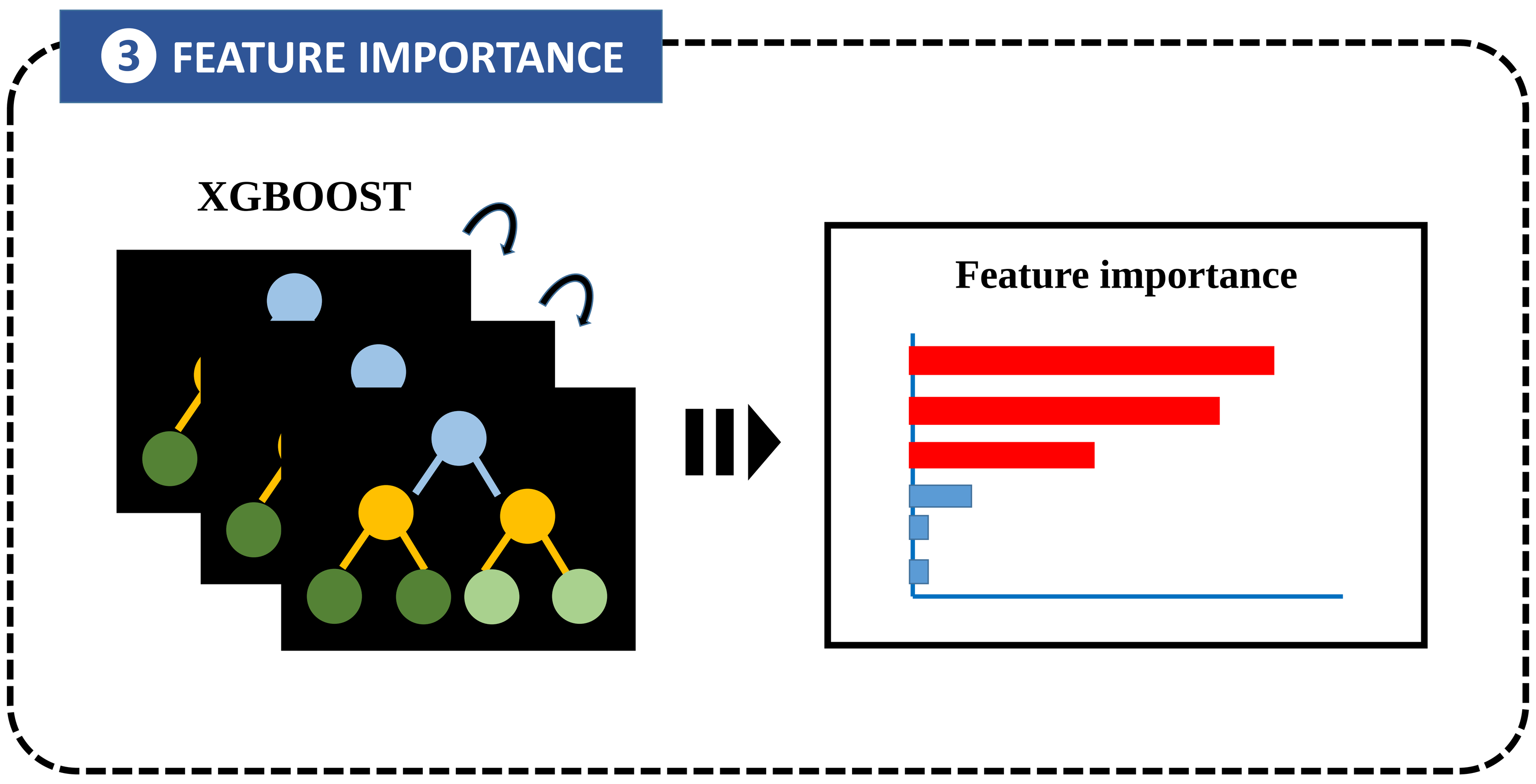
Abstract

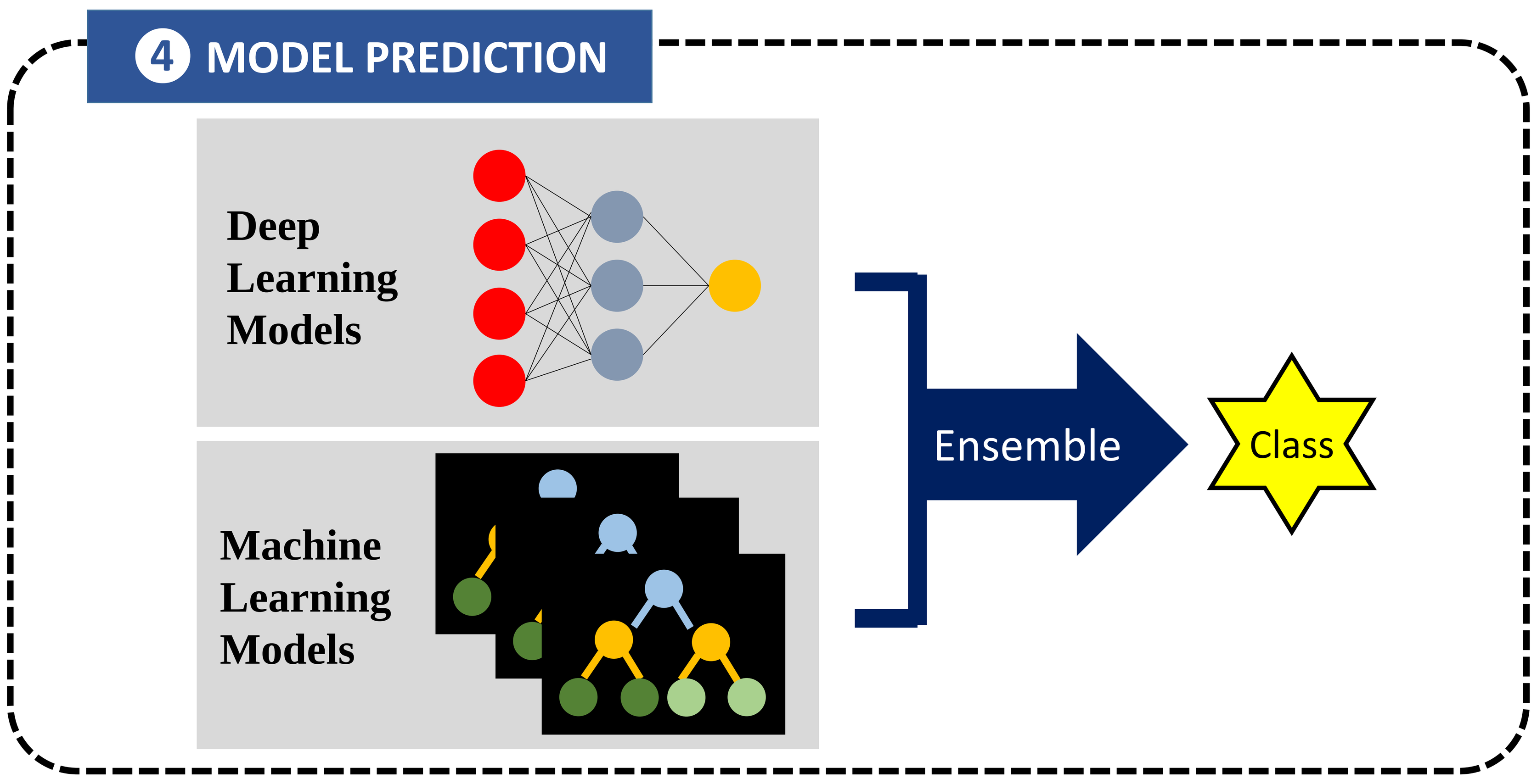
Cellphones have blended in our lives deeply nowadays. Since the public believe that people have distinct cellphone usage reaction between on duty and off working. National Health Research Institutes (NHRI) starts collecting many physicians' cellphone usage behavior as training data. Also, utilizing GPS to locate and simply determine whether they work at that time or not as training label. Then, through individual cellphone usage habit, we can analyze people's working status. Namely, we can achieve our goal knowing people's working efficiency on duty. In this research, we will employ machine learning method to obtain crucial features first. Next, using deep learning models and machine learning models to train on those selected features. Finally, we ensemble results of these models for prediction and analysis.

Method

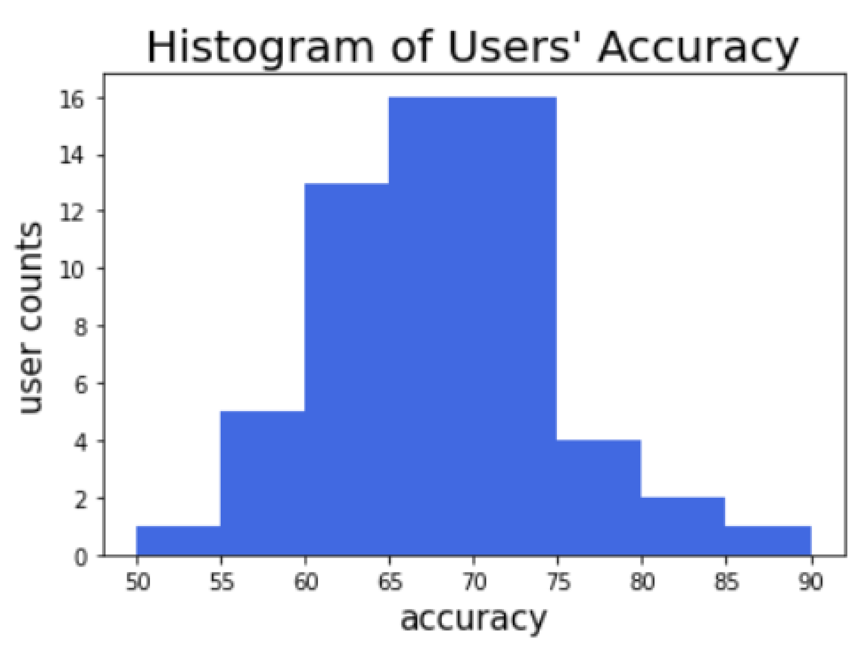




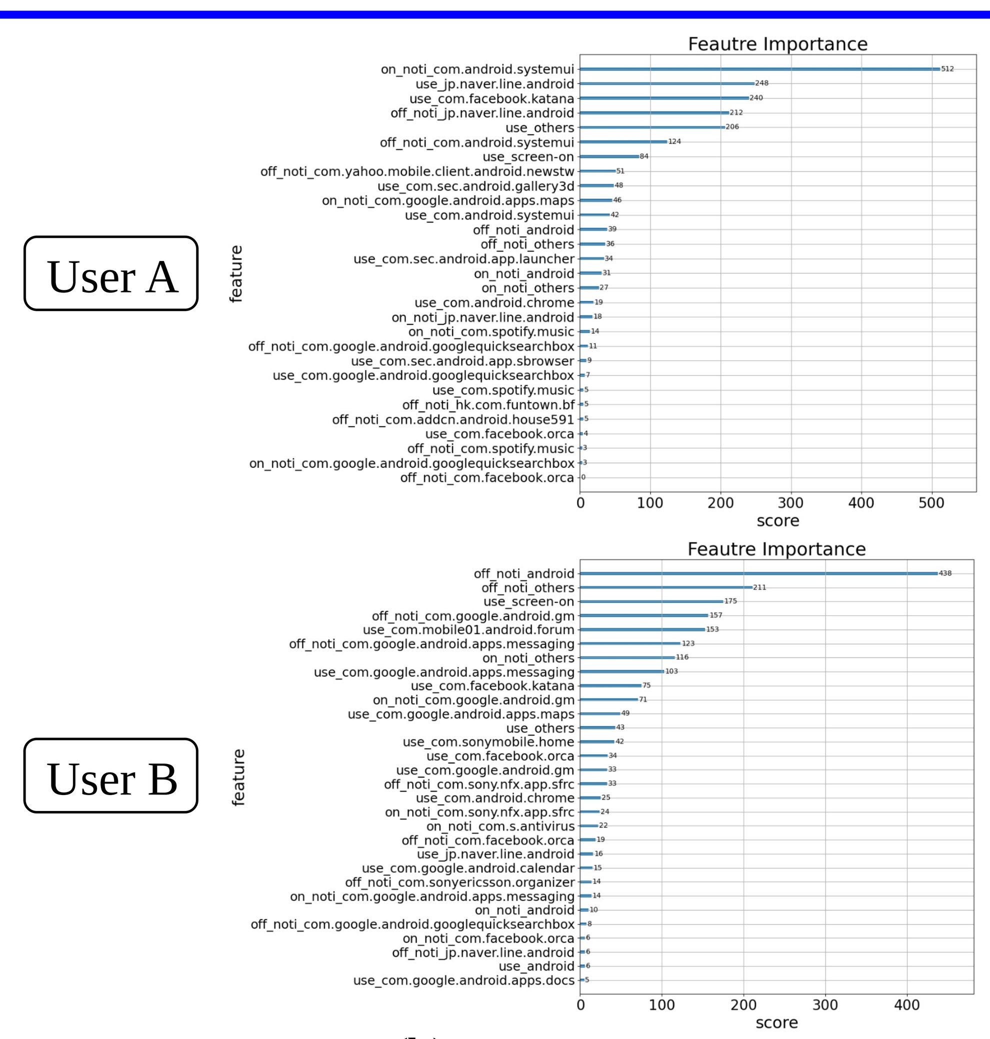




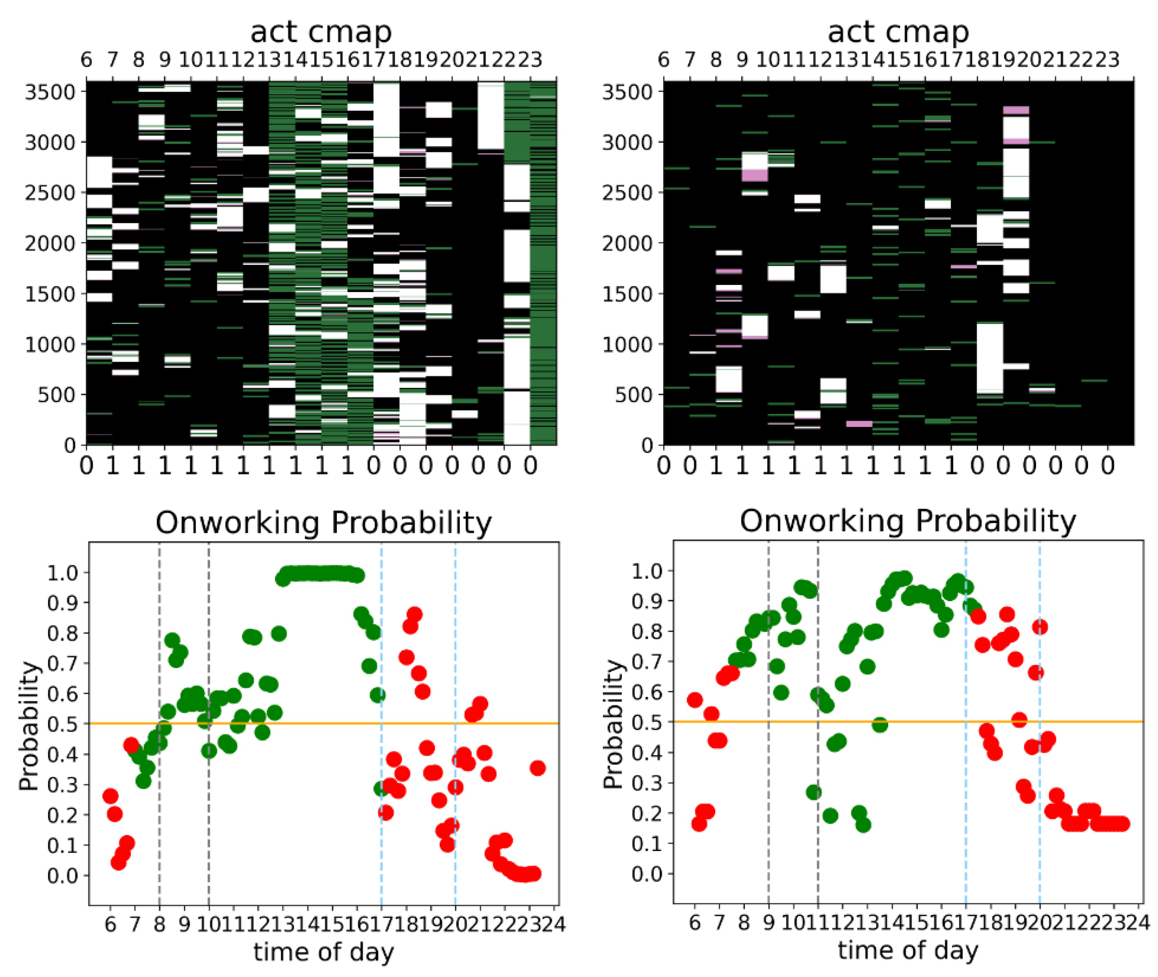
Results



(a) Accuracy Histogram



(b) Feature Importance



(c) Visualization – Working Probability

Conclusions

From the accuracy distribution, we can find that accuracy is not very high, since there are confused data which values in all features are similar but the label is different in the dataset. For Example, sometimes we don't use cellphone on duty. Also, we don't use it off working occasionally. Hence, in this situation, these data will become confused data. This is a vital issue that we can focus on in the future. Then, in the important feature graph, it shows that different users have distinct app usage habits, so this is a reason why we need to analyze it individually. Also, from visualization graphs, we can find that the behavior of using the cellphone is sometimes more like off working status during working time. It is a high chance that the user had lower working efficiency and by working probability plots, we can also obtain more information about working efficiency in detail.

References

[1] J. Ye, J.-H. Chow, J. Chen, and Z. Zheng. Stochastic gradient boosted distributed decision trees. In Proceedings of the 18th ACM Conference on Information and Knowledge Management, CIKM '09
[2] O. Chapelle and Y. Chang. Yahoo! Learning to Rank Challenge Overview. Journal of Machine Learning Research - W & CP, 14:1–24, 2011.
[3] L. Breiman. Random forests. Maching Learning, 45(1):5–32, Oct. 2001.
[4] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. JMLR, 2014.