题目：黑色星期五销售数据分析及用户画像、购买力预测

成员：雷云霖 何祎君 徐琳锴

摘要：黑色星期五在美国是一个购物狂潮期。在这个时期，各个商店会进行一次大规模的促销，从而引起群众大量购买物资。在购物盛典中将会产生大量的销售数据，这些数据具有非常大的商业价值，对其的挖掘可以给商家提供非常有效的建议让其提高销售额。我们尝试分析了Analytics Vidhya公司公布的2018年黑色星期五部分销售数据，用Apriori算法挖掘商品的关联规则，以此给商家商品捆绑销售提供建议。用随机森林模型用年龄、性别、职业、定居时间、居住城市、婚姻状态等属性对用户购买力进行预测，以此给商店为老顾客促销提供建议，并用线性回归模型进行比较，在十折交叉验证的框架下使用RMSE衡量模型。用随机森林、KNN、朴素贝叶斯和决策树等多种模型用购买记录预测用户性别、年龄，达到部分用户画像的功能，给精准营销提供建议，并比较这些模型的优劣，使用准确率衡量模型。我们对结果进行了合理的可视化表示，方便商户通过结果进行决策。

参考文献列表：

[1] Agrawal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB* (Vol. 1215, pp. 487-499).

[2] Zhang, Y., Pang, L., Shi, L., & Wang, B. (2014). Large scale purchase prediction with historical user actions on B2C online retail platform. *arXiv preprint arXiv:1408.6515*.

[3] Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997). Dynamic itemset counting and implication rules for market basket data. *Acm Sigmod Record*, *26*(2), 255-264.

[4] Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

[5] Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, *2*(3), 18-22.

[6] Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, *13*(1), 21-27.

[7] Breiman, L. (2017). *Classification and regression trees*. Routledge.

[8] Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, *1*(1), 81-106.

[9] Quinlan, J. R. (2014). *C4. 5: programs for machine learning*. Elsevier.

[10] Jia, R., Li, R., Yu, M., & Wang, S. (2017, July). E-commerce purchase prediction approach by user behavior data. In *Computer, Information and Telecommunication Systems (CITS), 2017 International Conference on* (pp. 1-5). IEEE.