Good morning, distinguished judges and professors. We are from Tsinghua High School. Our paper’s topic is Evaluation and Prediction of Cell Phone Sales Based on Various Methods.

This is the agenda of our presentation. We will mainly discuss the following eight sections.

To begin with, we will introduce the research background, purpose, and general process. With technology advancement and modern fast-pace life, more and more people prefer online shopping to directly going to shopping centers. For instance, Amazon, Alibaba are all familiar shopping websites to us. From manufacturers and sellers’ point of view, what matters the most is the key factors to sales promotion. Therefore, our research is mainly focused on this topic: what can promote sales and how to promote sales. Taking the feasibility of acquiring data into consideration, we secure our target to cellphones after evaluating various commodities.

Before conducting our own researches, we first conduct literature review. Previous researches mainly utilize the following three methods. For example, Fatao Wang employs Grey Relational Analysis to determine the Impact Factor of online shopping. Youzhi Xue employs C2C Model to find the relation between price and sales volume. Yanli Ma employs BP Neural Network Fitting to construct an e-commerce credit index evaluation system. However, there are some weaknesses. One is that they fail to analyze impact factors in a systematic and comprehensive way. The other is that they fail to reveal specific characteristics that contribute to higher sales volume. Therefore, our research initiates from filling in these academic blanks. To be more specific, our research purposes consist of three parts: first we rank the influential factors of sales volume; then we determine the specific characteristics that contribute to high sales volume; lastly, we predict sales volume of cell phones with given characteristics with our model.

This is the general process of our research. After acquiring sales records from Ali Express, we first extract useful information into relevant variables. Then we continue processing the data and reducing the number of variables without the loss of the core information in order to simplify for further modeling. Next, we construct our model with various methods and manage to rank main factors that influence sales volume and determine the influence of single variable to sales volume. After that, we optimize each model to determine the specific characteristics that contribute to high sales volume and discover the influence of multivariable to sales volume. Finally, we predict cell phone sales volume using our model and conduct sensitivity analysis and reach conclusion.

Then we come to the second part of today’s presentation. We will introduce basic assumptions made to simplify real-world implications without the loss of the core. Since our research mainly focuses on influential factors to sales volume, and the raw data fail to include information about sales volume, how to represent this concept using other index is the key problem. The raw data we acquire include these two indexes: conversion rate and click rate. According to economic definition, conversion rate equals valuable clicks divided by clicks, and click rate equals clicks divided by views. Therefore, it is natural to reach the following equation. We can employ click rate and conversion rate to reflect sales volume.

Then we come to the application. In general, we predict cell phone sales volume with our model. We classify those with only one blank parameter as the test sets, and employ these data to our model constructed by XG Boosting algorithm. For example, in this table, each row is a cell phone. And the first few columns are different variables, or cell phone parameters, like RAM, CPU, and these are just examples listed here. We have in total 25 variables that are not blank. Then our model manages to predict the conversion rate and click rate of each cell phone with the given characteristics. When compared with the real conversion rate and click rate, it’s clear that our results are very close to the true value. We also cite this formula to grade our results. And we receive an average score of 9.81 for click rate and 9.74 for conversion rate indicating that our model has high application value and accuracy.

Then we come to the sensitivity analysis to assess the accuracy and stability of our model. As we all know, sensitivity analysis is to detect the stability of our model, especially when the given data are inaccurate. In order to have more appropriate results, we divide all the variables into two categories: continuous variables and discrete variables. For continuous variables, we find that if we increase or decrease the original data with 1%, our three main models respond with little effect. The results from Principal Component Regression and BP neural network will correspondingly fluctuate within 1%, while the results from Bayes Distinction barely change. As for discrete variables, we change the value of only one parameter to adjacent category each time, leaving other parameters unchanged. For example, if we examine the stability of ROM, we change 2G to 4G, 4G to 8G, etc. and if the phone has 64G ROM, which is the largest, then we remain the data unchanged. We find that the output of the date changes approximately 0.1%. From the results, we can reach the conclusion that our model has universality and can be applied to more situations. For instance, if there are some errors in the data, our results will not vary rapidly in correspondence. Therefore, our model is relatively stable.

conclusion:

After all the modeling process, we finally reach conclusions.

For the first purpose of our research, that is to rank influential factors, we achieve this goal mainly through data extracting. So we synthesize the results from information entropy algorithm and PCA to reach the following ranking. From this ranking, we can know the influence and importance of each variable, and we can therefore choose the top few variables for further modeling in order to simplify without the loss of the core information.

As for the modeling process, we first yield qualitative results based on the weight determination technique. They can be depicted in such diagrams in which we can easily draw conclusions. Take Display Resolution for instance. Higher group number indicates higher sales volume, that is click rate 5 has the highest sales volume. So we can see that phones in Display Resolution Group 3 possess the greatest proportion of click rate 5, which means that phones with medium display resolution sell the best. Similarly, medium display resolution, low or high recording resolution, high camera resolution, and medium price are the most popular traits.

As for the quantitative conclusions, we determine a few specific cell phone traits that contribute to high sales volume. For example, according to our model, a perfect phone with high sales volume would have medium battery capacity, gold or white color, and high camera resolution.

In addition, from the manufacturers’ prospect, RAM, ROM and CPU should be the top choices when improving cell phone's overall quality. Besides, the public are not sensitive to changes in Display and Camera Resolution, so manufacturers should give more thought to lowering the price when these two factors are taken into consideration.

Last but not least, we will introduce the advantages of our model. To begin with, our research can help manufacturers understand public demands more precisely, and based on the demands, manufacturers can be clear about how to improve specific cell phone traits. Most ground-breakingly, our model can predict cell phone sales with given traits. In this way, it can assist manufacturers and sellers with insightful analysis and help them gain maximum benefits. Therefore, our model has high application value in the economic field.