NITTE EDUCATION TRUST | N.M.A.M. INSTITUTE OF TECHNOLOGY
(An Autonomous Institution affiliated to Visvesvaraya Technological University, Belagavi)
Nitte – 574 110, Karnataka, India
(ISO 9001:2015 Certified), Accredited with 'A' Grade by NAAC
☎: 08258 - 281039 – 281263, Fax: 08258 – 281265
Department of Computer Science and Engineering
B.E. CSE Program Accredited by NBA, New Delhi from 1-7-2018 to 30-6-2021

Mini Project Synopsis

Project Title

"MOBILE PHONE PRICE PREDICTION USING NAIVE BAYES ALGORITHM"

Submitted by:

4NM16CS122 Sahana Kamath

4NM16CS149 Sowjanya

Submitted to:

Ms. Divya Jennifer D'Souza

Assistant Professor Gd. I

Department of Computer Science and Engineering

2019

TABLE OF CONTENTS

ABSTRACT

In this project, main intention is to explore and analyze a data set which contains the specification of two thousand mobile phones and try to predict optimum price ranges for a list of mobile phones in the market by applying Naive Bayes Machine Learning Algorithm.

LITERATURE SURVEY

Using previous data to predict price of available and new launching product is an interesting research background for machine learning researchers. Sameerchand-Pudaruth predict the prices of second hand cars in Mauritius. He implemented many techniques like multiple linear regression, KNN, Decision Tree, and Naïve Bayes to predict the prices and got Comparable results from all these techniques.

Shonda Kuiper used multivariate regression model to predict price of 2005 General Motor cars. He collected the data from www.pakwheels.com. The main part of this research work is Introduction of suitable variable selection techniques, which helped to find that which variables are more suitable and relevant for inclusion in model. This helps students and future researchers in many fields to understand the conditions under which studies should be conducted and gives them the knowledge to discern when appropriate techniques should be used.

METHODOLOGY

NAIVE BAYES CLASSIFIER :

Naive Bayes classifiers are linear classifiers that are known for being simple yet very efficient. The probabilistic model of naive Bayes classifiers is based on Bayes' theorem, and the adjective naive comes from the assumption that the features in a data set are mutually independent.

In probability theory and statistics, Bayes' theorem (alternatively Bayes' law or Bayes' rule) describes the probability of an event, based on prior knowledge of conditions that might be related to the event.

Bayes' theorem is stated mathematically as the following equation.

$$P(B) = \frac{P(A)P(A)}{P(B)}$$

where A and B are events and P(B) $\neq$ 0.

- $P(B)$ is the conditional probability : the likelihood of event A occurring given that B is true.

- $P(A)$ is also a conditional probability : the likelihood of event B occurring given A is true.
- $P(A)$ and $P(B)$ are the probabilities of observing A and B independently of each other; this is known as marginal probability.

DATASET :

The dataset for this problem is collected from Kaggle which is available under Public Domain License. The dataset consists of 2000 samples of mobile phones which specifies its features and price range.

We are going to use the following variables (or features) to train our Naive Bayes classifier

1. Id : ID
2. battery_power : Total energy a battery can store in one time measured in mAh
3. blue : Has bluetooth or not
4. clock_speed : speed at which microprocessor executes instructions
5. dual_sim : Has dual sim support or not
6. fc : Front Camera mega pixels
7. four_g : Has 4G or not
8. int_memory : Internal Memory in Gigabytes
9. m_dep : Mobile Depth in cm
10. mobile_wt : Weight of mobile phone
11. n_cores : Number of cores of processor
12. pc : Primary Camera mega pixels
13. px_height : Pixel Resolution Height
14. px_width : Pixel Resolution Width
15. ram : Random Access Memory in Megabytes
16. sc_h : Screen Height of mobile in cm
17. sc_w : Screen Width of mobile in cm
18. talk_time : longest time that a single battery charge will last when you are
19. three_g : Has 3G or not
20. touch_screen : Has touch screen or not
21. wifi : Has wifi or no

PROCEDURE :

1.DATA COLLECTION :

We collected the data set Kaggle which is available under Public Domain License. The dataset The data set consists of 2000 samples.

2.DATA SET CATEGORY:

The data set is divided into test data and train data based on the spec ratio of 0.67.

Here we have assumed four price ranges 0,1,2,3.

0 - Price below 10,000

1 - Price below 20,000

2 - Price below 30,000

3 - Price below 40,000

We have used Python language to predict the price range of mobile phones by using Naive Bayes's theroem.

OBSERVATION AND DISCUSSION

For our experiments, we train the Naive Bayesian Model for the 'price prediction' data set to predict the range of test data accurately.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	battery_po	blue	clock_spee	dual_sim	fc	int_memor	m_dep	n_cores	pc	px_height	px_width	ram	touch_scre	wifi	price_range
2	842	0	2.2	0	1	7	0.6	2	2	20	756	2549	0	1	1
3	1021	1	0.5	1	0	53	0.7	3	6	905	1988	2631	1	0	2
4	563	1	0.5	1	2	41	0.9	5	6	1263	1716	2603	1	0	2
5	615	1	2.5	0	0	10	0.8	6	9	1216	1786	2769	0	0	2
6	1821	1	1.2	0	13	44	0.6	2	14	1208	1212	1411	1	0	1
7	1859	0	0.5	1	3	22	0.7	1	7	1004	1654	1067	0	0	1
8	1821	0	1.7	0	4	10	0.8	8	10	381	1018	3220	0	1	3
9	1954	0	0.5	1	0	24	0.8	4	0	512	1149	700	1	1	0
10	1445	1	0.5	0	0	53	0.7	7	14	386	836	1099	0	0	0
11	509	1	0.6	1	2	9	0.1	5	15	1137	1224	513	0	0	0
12	769	1	2.9	1	0	9	0.1	5	1	248	874	3946	0	0	3
13	1520	1	2.2	0	5	33	0.5	8	18	151	1005	3826	1	1	3
14	1815	0	2.8	0	2	33	0.6	4	17	607	748	1482	0	0	1
15	803	1	2.1	0	7	17	1	4	11	344	1440	2680	0	1	2
16	1866	0	0.5	0	13	52	0.7	1	17	356	563	373	0	1	0
17	775	0	1	0	3	46	0.7	2	16	862	1864	568	1	1	0
18	838	0	0.5	0	1	13	0.1	8	4	984	1850	3554	0	1	3
19	595	0	0.9	1	7	23	0.1	3	17	441	810	3752	1	0	3
20	1131	1	0.5	1	11	49	0.6	5	18	658	878	1835	1	0	1
21	682	1	0.5	0	4	19	1	4	11	902	1064	2337	1	1	1
22	772	0	1.1	1	12	39	0.8	7	14	1314	1854	2819	1	0	3
23	1709	1	2.1	0	1	13	1	2	2	974	1385	3283	0	0	3

Fig 1 - Mobile phone specification data set

We have used Naive Bayes theorem with a split ratio of 0.67. That is out of 2000 rows we have taken 1340 rows of data as training data and 660 rows of data as test data.

RESULT

The following shows the figure accuracy obtained while classifying our test data into our implementation Naive Bayes Algorithm.

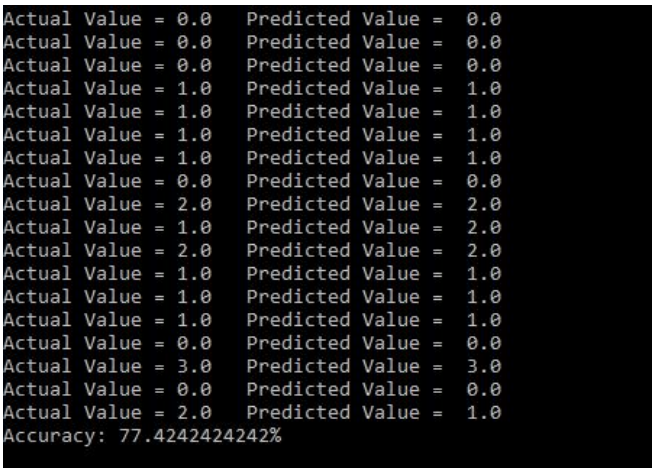

Fig 2 - Output of experiment

We have obtained an accuracy of 77.42%.

CONCLUSION

In this project we have successfully implemented the Naive Bayes algorithm for predicting mobile price range. This work can be used in any type of marketing and business to find optimal product (with minimum cost and maximum features).This kind of prediction will help companies estimate price of mobiles to give tough competion to other mobile manufacturer. Also it will be useful for Consumers to verify that they are paying best price for a mobile.

REFERENCES

1. DATASET : https://www.kaggle.com/vikramb/mobile-price-prediction

2. https://github.com/vikram-bhati, Classification classify mobile price range