FEATURE EXTRACTION FOR MOBILE HANDSET IN COHERENCY WITH PRICING FACTORS

Anurag Tiwari\textsuperscript{1}, Vivek Kumar Singh\textsuperscript{2}, Praveen Kumar Shukla\textsuperscript{3}, Manuj Darbari

\textsuperscript{1}Research Scholar, Computer Science, BBD University, Lucknow, Uttar Pradesh, India.
\textsuperscript{2}Professor, Department of Computer Science, BBDNIIT, Lucknow, Uttar Pradesh, India.
\textsuperscript{3}Professor, Department of Information Technology, BBDNIIT, Lucknow, Uttar Pradesh, India.
\textsuperscript{4}Professor, Department of Information Technology, BBDNITM, Lucknow, Uttar Pradesh, India.

Email: \textsuperscript{1}anuragrtiwari@gmail.com, \textsuperscript{2}viveksinghbbd@gmail.com, \textsuperscript{3}drpraveenkumarshukla@gmail.com, \textsuperscript{4}manujuma@gmail.com

Corresponding Author: Anurag Tiwari

https://doi.org/10.26782/jmcms.2020.03.00024

Abstract

This paper presents a showcase of analysis of Mobile price with respect to the features it is able to analyse for the buyer. The paper gives machine learning approach in identification of the right price and its subsequent features detail. ANN with Back propagation algorithm has been chosen by developing a customized mobile selection algorithm using Kaggle database for modelling and Analysis. Various cost factors are adjusted in relation with the features to be incorporated in the Handset. The adjustment of input variables is done by the help of the machine learning technique giving the exact relationship in three main factors Requirement of the customer based on their segmentation, Price and Features.

Keywords : Mobile Selection Criteria, Machine Learning, ANN, DSS.

I. Introduction

The Customer of prospective mobile phone facing the problem related to selection of best handset based on the preference in his mind with the budget he has “He would buy something best in optimum price range”. Multi-objective is a field of computer science which enable the system to answer such type of predictive price (with the help of machine learning according to tech ARC’s Key trends. The report
shows dip in 2% sales of Smartphone by 2020 due to falling sales of features phone or the user of particular segment not setting the proper phone feature as in his budget. Now the customer & company both are facing this innovation is issue with what features are essential at what price segment or we can say the optimization of phone cost with available feature as per customer preference.

Moving very fast now new mobile with diversified features are launched; Mobile industry is changing industry everyday new features are added to cost of the phone but finding better (Optimum product) is the concern from the user prospective.

Based on the preference some problem also eats the real start motor vehicle purchase food item medicine health supplement production etc. The feature & its preference is a very important to estimate the price of handset [V]. For example battery backup, screen size, connectivity, processor etc. The important aspect of the mobile phone in 21st century as many kind of processor exits in which every customer what to buy best but it effects the price, some while considering the screen size the senior citizen mostly prefer the big screen, house wife’s prefer High MP camera likewise student gives priority to the processor because they love to play game. Based on the list of feature the price of phone varies based as the feature & priority the system recommend the mobile phone & price rang to the user.

Our problem domain revolves round buying optimization from customer’s point of view where the customer has to make a choice of various handset available online on a particular e marketing website shown in fig 1.

![Fig. 1: A comparative Analysis of the Handset price Vs features[XXIII]](image)

The customers are classified as business customer, Young generation and Housewife this is to be related like

i) RAM

ii) Camera front and Back

*Copyright reserved © J. Mech. Cont. & Math. Sci. Anurag Tiwari et al*
iii) Battery Life  
iv) Processing Capability  
v) Cost consideration  
A mapping has to be established between three Customer group segment and four features as shown in the Table I.

**Table I: A comparative Analysis of Feature Vs Customer segment**

| Customer Segment | Features       |   |   |   |   |
|------------------|----------------|---|---|---|---|
|                  | RAM | Camera | Battery | Processing | Cost |
| Housewife        | -   | ×      | ×        | ×           | ×    |
| Young User       | ×   | -      | ×        | ×           | ÷    |
| Business         | ×   | ÷      | ×        | ×           | ÷    |

i) Requirement level $\times$ = High  
ii) Requirement level $\rightarrow$ $\circ$ Moderate  
iii) Requirement level $\rightarrow$ $\rightarrow$ May be Required May be NOT  

We can speculate the mapping using “Requirement Level” Ranging from High to May/May NOT level. Requirement Structure should be decided but their requirements mapping are not so accurate, hence a optimization technique is required. Optimization has played significant role in our life, be it financial decision making or Handset selection. There are many criteria for decision making suggested till date ranging from simple optimization technique to highly Multi-dimensional decision making. Earlier their optimization technique was Gradient Descent algorithm, but it has got certain limitation. We have tried to overcome this shortcoming by using machine learning.

**II. Related Work**

Some prominent work related to pricing and feature strategy selection using Machine Learning which have motivated the authors in developing a model for mobile handset.

The random forest [II] is methodology to predict the expected price of the used car based on the feature selected by user and SameerchandPudaruth [XII] using the existing data set to guess the price of an product is very useful background for machine learning[XXIX].He used many technique like multiple linear Regression (KNN), Decision tree & Naive Bayes to predict the prices & also analyzed the companion his model has been motivating factor for us to implemented it for Mobile handsets[VIII],[IX],[XI]. On the similar lines Shonda Kuiper [XII] has done his research for the GM Cars to predict the price by using multivariable Regression model in this he collected data from the online website.
www.pakwheets.com introducing a variable selection Technique to find the variable that is more perfect.

Mariana Listiani[XIII] used support vector machine for the same work & it is found that this technique is more suitable to find the result as compare with other existing at that time like multiple linear regressionspecially case of large dataset Limsombunchi [XIV] predicting the estimated price of the home in his research work. His work has better result as compare to Hedonic method.

Another research by Bourassa et al [XXIV] shows how the price of a house associated with the price of immediate house. In this research total 4 model were used and each model has their own difference two ordinary least sequences (OLS) four geo-statistical & two lattice model which show how to apply sub variable models, they able to provide higher guessing accuracy of OLS having sub market variable which has better accuracy then others. The value of percentage reached to 46.8 % from 39.8 with adding up submarket variable to OLS model. Later on Nau [XXV] in his research simple regression model were applied and a linear bonding exists Z=Bi(Xi) variable Xi and outcome values Z where the B is constant where variance in the variable z & variable y is nearly same and anonymous variation in Zwhich is anonymous, and independent and of randomized. Gonggi [XXIX] in his research proposed a model that used ANN for predicting the cost of car it focuses on non-linear relationship in data that was used for better yield as compare to linear relationship models.

Lei Dong [XXX] et al in his research work used the features extracted from the data set (Restaurant), training with machine learning model to determine day time and night time population and analysis trade-off between accuracy, special resolution and no of training sample[XVI].

Sundsøy el al [XXXII] implemented in his research work a simple deep learning model & given a comparison with a simple data mining and to show, how socio economy status in large defines identify of mobile data set categorizing phone usage, topup pattern, handset type, social network structure & individual mobility achieving 72% AVL on testing, we have automated the process of mapping both the domains using various optimization Techniques. There are various algorithms available but the trends is shifting to Machine learning for accurate data analysis, since our problem deals with predicting consequents behavior we have to go about ANN approximation based strategy[XXIV]. K Noor and Saddaqat J [XV] also worked to predict the price of Vehicles using different techniques. The researchers achieved highest accuracy using multiple linear regression[XIX]. This paper proposes a system where price is dependent variable which is predicted, and this price is derived from factors like vehicle’s model, make, city, version, color, mileage, alloy rims and power steering[IV].Namhyoung [XXXI] worked on model performance & gradient boosting mechanism on train test data 75/25 where the base classifiers are built sequentially. The algorithm combines the new classifier with ones from previous iterations in an attempt to reduce the overall error rate[XX]. We believe that though identifying the complete process may not be possible, we can still detect certain patterns or regularities[XXVIII]. Assuming that the irregularities are maintain in
buying behavior we analyzed it in terms of an Algorithm which can learn the buying behavior and give an insight to the company that what are the main parts and how it will impact on the cost of the final product[XXI]. Machine learning is not just a database problem, it is a mix of Artificial Intelligence, to optimize a performance criterion using example of data set[XXII],[XXVI].

III. Machine Learning Algorithm used in Analysis

For our Analysis we have used the concept of Multilayer Perception feed forward [III] with a Softmax-output Activation function. The training in this is achieved by the help of Back Propagation Algorithm with Gradient Decent learning rate. The training stops when error reaches to Zero value. The training is given by implementing the chain rule given as:

\[
\frac{dE}{d\text{net}_j} = \frac{dE}{d\text{o}_j} \cdot \frac{d\text{o}_j}{d\text{net}_j} \tag{I}
\]

After calculating the final derivation function. We obtain the

\[
\frac{dE}{d\text{net}_j} = -(t_j - o_j) o_j(1 - o_j)
\]

Which is a derivate of sigmoid function where

\[
\text{net}_j = \sum w_{j_i} x_{j_i} \text{ (The weight sum of in for unit } j) \\
\]

\[
t_j = \text{target output for unit } "j" \\
\]

\[
o_j = \text{Output computed by unit } "j" \\
\]

\[
E_d = \text{error on training case of } "d"
\]

A complete pass over the entire pattern in the training set is called an ep.For number of outputs we have algorithm given as

\[
\text{initialise } v_{h} \text{ and } w_{h} \\
\]

Repeat

For all \((x', r') \in x \text{ in random order})

For \(h = 1..h\)

\[
z_h = \text{Softmax}(W_{x}^r, x^r)
\]

For \(i = 1 ... k\)

\[
y_{i} = v_{i}^T z
\]

For \(h = 1..h\)

\[
\Delta w_{h} = \left( \sum_{i} (y_{i}' - y_{i}) v_{i} \right) z_{h}(1 - z_{h})
\]

For \(h = 1 ... H\)
IV. Problem Analysis

Our problem is mainly focused on the following factors:

- Battery – power
- Clock speed
- Dual sim feature
- Internal Memory
- Mobile Weight
- No of Core
- Height
- Width
- RAM
- Talk Time
- Generation Supported
- Touch Screen Size

We will be analyzing the results of the parameters with respect to company and its impact on consumers.[XVII]. Our analysis focuses on finding out the impact of all these actors using the concept of Machine learning. We have used the concept of Artificial Neural learning [V], [XVIII]. We have used the concept of Artificial Neural Network [XXVII], [IV] using Anaconda Interface with coding done in Python language. We have used above parameters to generate different plots to analyses our result. The Table II present the Data taken from KaggleDatabase.

Table II: Various Variables used in Mobile Handsets

| Index | Battery Power | Clock Speed | Dual Sim | RAM | Touch Screen Size | Mic | RAM Type |
|-------|---------------|-------------|----------|-----|------------------|-----|----------|
| 1     | 1.0           | 2.0         | 3.0      | 4.0 | 5.0               | 6.0 | 7.0       |
| 2     | 0.5           | 1.5         | 2.5      | 3.5 | 4.5               | 5.5 | 6.5       |
| 3     | 2.0           | 3.0         | 4.0      | 5.0 | 6.0               | 7.0 | 8.0       |

Copyright reserved © J. Mech. Cont. & Math. Sci.

Anurag Tiwari et al
On the basis of the above data we compare it with various brands to see the correlation and what the other brands should focus to improve in order to increase to reach of their brand.[XVIII][XIX]. It is done by the help of Perceptual Mapping graph showing the product placement of Various Brands vs Cost. (figure 2). Manufacturing companies use perceptual mapping to compare products (and potential products) based on the weightage given by customers.

![Perceptual Mapping Graph]

**Fig. 2:** Perceptual mapping of Various Brands of Mobile Manufacturer

The main idea of a perceptual map is identification of images that consumers have of and the reactions they have to brands, products, services and other market offerings, like in this case brands like Apple, Nokia, Samsung, Redmi, Oppo, Realme and Micromax are compared and are positioned in matrix based on feature and cost, now the question is how can Samsung migrate to far better position as that of Apple but still compromising on Cost factor to increase its Reach, for this we have taken Machine Learning as one of the handy tool to interpret solution to this problem. Based on the above graph we have developed a Training Algorithm (figure 3)
train ANN($f, w, o, \theta$)
For epochs from 1 to N
   While ($j = m$)
      Randomly initialize $w = \{w_1, w_2, ..., w_n\}$
      input $o = \{o_1, o_2, ..., o_m\}$ in the input layer
      forward propagate ($f'$) through layers until getting the predicted result $y$
      compute $e = y - y'\$
      back propagate $e$ from right to left through layers
      update $w$
   end loop
end loop

$f = \{f_1, f_2, ..., f_i\}$ where $n$ is number of features of mobile handset.
$W = \{w_1, w_2, ..., w_n\}$ where each $w_i$ correspond to weight given to node $i$ for feature $f_i$
$o = \{o_1, o_2, ..., o_n\}$ where $m$ is number of observations.
$E$ is error obtained from subtracting predicted result $y$ from actual result $y'$.

$f' = w_i$ propagate through each neuron & activate it if the result of product is greater than the defined thresholds.

**Fig. 3:** Algorithm for Mobile Handset selection Based on Various features

**V. Statistical Analysis for Strength Prediction**

The strengthening of concrete is a complex process involving many external factors. A number of improved prediction techniques have been proposed by including empirical or computational modeling, statistical techniques. Many attempts

The formal application of the above algorithm is applied on Spyder and coded in Python 3.7 IDE to get the following outcomes is the form of various plots. A plot between RAM and Price is being developed shown in figure 4 using ANN linear Regression output.

![Data Analysis Graph showing Relation between RAM and Price](image_url)

**Fig. 4:** Data Analysis Graph showing Relation between RAM and Price
In this field of Machine learning the problem of statistical classification, a confusion matrix also known as Error Matrix is a specific table layout that allows visualization of the performance algorithm, typically a supervised learning one. Table III of the precision shown Recall f1 score and support plotted for Confusion Matrix.

A Confusion matrix is a summary of prediction result on a classification problem. The number of prediction depending on its validity is classified by the help of enumeration known as Count Values and broken down by each class. The confusion matrix shows the ways in which our model can be divided to make predictions.

It gives an in eight of the errors being made by a classifies and the type of the error also.

**Table III** : Precision, Recall, f1_score and Support Values derived from Kaggle

|                  | Class 1 Predicted | Class 2 Predicted |
|------------------|-------------------|-------------------|
| Class 1 Actual   | Time Positive     | False Negative    |
| Class 2 Actual   | False Positive    | Time Negative     |

**Class 1** is defined as Positive

**Class 2** is defined as Negative

where:

Positive (P): Observation is positive
Negative (N): Observation is not positive
Time Positive: Observation and prediction positive
False Negative: Observation is positive but prediction negative.
Time Negative: Observation is Negative and is predicted to be negative
False Positive: Observation is negative but prediction positive

| precision | recall | f1-score | support |
|-----------|--------|----------|---------|
| 0         | 0.99   | 0.98     | 0.98    | 95      |
| 1         | 0.95   | 0.96     | 0.97    | 92      |
| 2         | 0.95   | 0.94     | 0.93    | 99      |
| 3         | 0.95   | 0.96     | 0.96    | 114     |
| avg / total| 0.96   | 0.96     | 0.96    | 400     |

A 99% accuracy can be excellent, good medicine, poor or terrible depending upon the problem, where as a High Recall indicate the clean is correctly recognized given by the formula as:
Recall = $\frac{TP}{TP + FN}$

And finally we get F measure as:

$$F_{measure} = \frac{2 \cdot Recall \cdot Precision}{Recall + Precision}$$

Each row of the matrix represent the instances in predicted class while each column represents the instances in actual class(or vice versa ). The name stress from the fact that it makes it easy to see if the system is confusing two classes(i.e. commonly mislabeling one as another).

It is special kind of contingency table (fig 5.) with two dimensions (“Actual and Predicted”) and identical sets of “Classes” in both dimensions (each combination of dimension and class is a variable in the contingency table)

**Fig. 5:** Confusion Matrix representing instances of the predicted class

![Confusion Matrix](image1)

**Fig. 6:** Accuracy of Train & Test

During the training of a machine learning model, the current state of the model at each step of the training algorithm can be evaluated. The evaluation is done by training dataset to give an idea of how well the model is “learning.” It can also be Tested on hold-out validation dataset that is not part of the training dataset. The final
outcome of the model is also tested by giving train and test data for 10 epoch 96% accuracy at 8th epoch (figure6).

![Fig.7: Accuracy Comparative Analysis of various Schemes](image)

From figure 7 we are able to confirm that ANN is best possible optimized solution set for our problem with about almost 97% accuracy.

**VII. Conclusion**

The paper highlights the uses of Machine learning Algorithm in Mobile Handset feature selection by comparing it with various price range. ANN is found to be most suitable for this type of Analysis by developing Basic propagation algorithm paper optimizes in algorithm to select various features under best price range. The Error Back propagation for layered feed forward network provides accurate result in the form of confession Matrix with diagnoses automatically shown be “Time positive” showing the system has achieved High Value of Accuracy in each at every segment.

**VIII. Acknowledgments**

I wish to record my deep sense of gratitude and profound thanks to my research supervisor Dr. Vivek Kumar Singh, Professor & Director, Department of Computer Science, BBDNIIT, Lucknow, for his keen interest, inspiring guidance, constant encouragement during all stages of Research work. I am grateful to my Co-Supervisor Dr. Praveen Kumar Shukla, Professor Department of Information Technology BBDNIIT for his valuable suggestions and support during the course of my research work. I also indebted to Dr. Manuj Darbari, Professor & Head, Department of Information Technology, BBDNITM, Lucknow for his valuable suggestions and great support during research work. I am greatly appreciating of my family members Spacialy My Father Mr. Ramkrishna Tiwari, who has encouraged me with his moral support to complete my research work.

*Copyright reserved © J. Mech. Cont. & Math. Sci.*

*Anurag Tiwari et al*
References

I. A Chaudhary, S. Kolhe and Rajkamal, “Performance Evaluation of feature selection methods for Mobile devices”, ISSN: 2248-9622, Vol. 3, Issue 6, NovDec 2013, pp.587-594.

II. A Lapedes and R. Farber, “How Neural Networks Works”, in Neural Information Processing Systems (D.Z. Anderson, ed.), (Denver), American Institute of Physics, New York, pp. 442-456, 1988.

III. Bourassa, S.C., Cantoni, E. and Hoesli, M. 2007. “Spatial dependence, housing submarkets, and house price prediction”, The Journal of Real Estate Finance and Economics, 35(2), p.143-160.

IV. GONGGI, S., 2011. New model for residual value prediction of used cars based on BP neural network and non-linear curve fit. In: Proceedings of the 3 rd IEEE International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)

V. H. Liu and R. Setiono, “A probabilistic approach to feature selection - A filter solution,” the 13th International Conference on Machine Learning, pp. 319-327, 1996.

VI. H. White, “Learning in Artificial Neural Networks: A Statistical Perspective”, Neural Computation, 1(4), pp.425-464, 1989.

VII. Kanwal Noor and Sadaqat Jan, “Vehicle Price Prediction System using Machine Learning Techniques”, International Journal of Computer Applications (0975 – 8887) Volume 167 – No.9, June 2017.

VIII. Khaidem, Luckyson & Saha, Snehanshu & Basak, Suryoday & Kar, Saibal & Dey, Sudeepa. (2016). Predicting the direction of stock market prices using random forest.

IX. Lei Dong, Carlo Ratti, Siqi Zheng, ‘Predicting neighborhoods’ socioeconomic attributes using restaurant data Proceedings of the National Academy of Sciences Jul 2019, 116 (31) 15447-15452;

X. Limsonbunchai, V. 2004. “House Price Prediction: Hedonic Price Model vs. Artificial Neural Network”, New Zealand Agricultural and Resource Economics Society Conference, New Zealand, pp. 25-26. 2004

XI. M. Hall, “Feature Selection for Discrete and Numeric Class Machine Learning”, Department of Computer Science.

XII. M. Robnik and I. Kononenko, “Theoretical and Empirical Analysis of ReliefF and RReliefF”, Machine Learning Journal, 2003.

XIII. M.C. Mozer,”A Focused Back – propagation Algorithm for Temporal Pattern Recognition”, Complex Systems, 3, pp.349-381, 1989.
XIV. Mariana Listiani , 2009. “Support Vector Regression Analysis for Price Prediction in a Car Leasing Application”. Master Thesis. Hamburg University of Technology.

XV. Minitab Express Support. Interpret all statistics and graphs for Multiple Regression.[Online] Available

XVI. Mobile data and specifications online available from https://www.flipkart.com/ (Last Accessed on Friday, December 22, 2019, 3:14:34 PM)

XVII. NamhyoungK, Kyu J, Yong.K, A New Ensemble Model for Efficient Churn Prediction in Mobile Telecommunication, 2012 45th Hawaii international conf on system sciences.

XVIII. Nau, R. 2014. Notes on linear regression analysis, Lecture handouts, Duke University, Furqa School of Business, 26 nov 2014.

XIX. Nisha Thomas and Mercy."Implementation of Back propagation Algorithm in Reconfigurable Hardware”. 2011.

XX. R.Linsker, “From Basic Network Principles to Neural Architecture”, in Processings of the National Academy of Sciences, 83,(USA),pp 7508-7512,8390-8394,8779-8783,1986.

XXI. S.Titri, H. Bourmeridja.”New Reuse Design Methodology for Artificial Neural Network Implementation”. 1999.

XXII. Sameerchand Pudaruth . “Predicting the Price of Used Cars using Machine Learning Techniques”. International Journal of Information & Computation Technology. ISSN 0974-2239 Volume 4, Number 7 (2014), pp. 753764

XXIV. Shonda Kuiper, “Introduction to Multiple Regression: How Much Is Your Car Worth? ”, Journal of Statistics Education · November 2008

XXV. Singh, Y., Bhatia, P. K., & Sangwan, O. 2007. “A review of studies on machine learning techniques”, International Journal of Computer Science and Security, 1(1), 70-84.

XXVI. SireeshaJasti ,TummalaSitaMahalakshmi”Multiple Linear Regression”, IJRTE, pp. 1919-1925, August 2019.

XXVII. Suebsing and N. Hiransakolwong, “Euclidean based Feature Selection for Network Intrusion Detection”, International Conference on Machine Learning and Computing IPCST, 2011.
XXVIII. Sundsøy, Pål & Bjelland, Johannes & Reme, Bjørn & Jahani, Eaman. (2016). Deep Learning Applied to Mobile Phone Data for Individual Income Classification. 10.2991/icaita-16.2016.24.

XXIX. Tadayoshi Horita, Takuroa Murata and Itsuo Takanami. “A Multiple Weight and Neuron Fault Tolerant Digital Multilayer Neural Network”. 2006.

XXX. Thu Zar Phyuy, Nyein Nyein Oo. Performance Comparison of Feature Selection Methods. MATEC Web of Conferences 42, (2016).

XXXI. X. Yao, “A New Evolutionary System for Evolving Artificial Neural Networks”, IEEE Trans. Neural Networks, 8, May 1997.

XXXII. Z. Karimi and M. Mansour and A. Harounabadi “Feature Ranking in Intrusion Detection Dataset using combination of filtering”, International Journal of Computer Applications, Vol. 78, September 2013.