Predicting Sales In E-commerce Using Bayesian Network Model

Wamukekhe Everlyne Nasambu[1], Waweru Mwangi [2], Michael Kimwele[3]

[1]Computing Department, Jomo Kenyatta University of Agriculture and Technology

P.O Box 62000 - 00200, Nairobi, Kenya

[2] Computing Department, Jomo Kenyatta University of Agriculture and Technology

P.O Box 62000 - 00200, Nairobi, Kenya

[3] Computing Department, Jomo Kenyatta University of Agriculture and Technology

P.O Box 62000 - 00200, Nairobi, Kenya

Abstract

An area of substantial research is that of predicting product sales, such as books, video games and movie tickets. There are a number of prediction models that have been used to predict future sales however these models attempt to solve the problem by making assumptions. These models assume that independent variables are truly independent. In theory, there should be zero correlation between any of the independent variables. In practice, however, many variables are related, sometimes quite highly. Therefore, different prediction techniques/methods have been and are being researched on and proposed

to address this drawback. The aim of this study is to identify ways of improving prediction of product sales in mobile phones. Consequently, the study will realize a predictive model that will classify sentiments from social media and compute the probability and present an improved predictive model.

Keyword: *Bayesian Networks, Graphical Models, Prediction Model, Social Media, Electronic Commerce*

1. Introduction

The expansion of the Internet in the past decade has given researchers new avenues to explore the art of prediction. First, online forums and blogs allowed individuals to share thoughts, Opinions and information with one another on any imaginable subject. In a research [1] examined the predictive ability of Internet chatter from user-generated content on blogs and forums and its impact on music album sales. It was concluded that that online chatter is predictive of album sales during the first two weeks following the album release and the week preceding the album release. A similar approach taken by [2] with movies and found correlation between references to movies in weblog posts both before and after their release and the movies' financial success. In a study to analyze the impact of sentiments analysis in prediction [3], it was concluded it had some impact especially after the movie had been released. Sentiments have become a pointer to stimulus in social media and a combination of sentiments and keywords increases the prediction accuracy.

2. BAYESIAN CLASSIFIER

Bayesian networks (BNs), also known as belief networks (or Bayes nets for short), belong to the family of probabilistic graphical models (GMs). These graphical structures are used to represent knowledge about an uncertain domain. In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables. These conditional dependencies in the graph are often estimated by using known statistical and computational methods. Hence, BNs combine principles from graph theory, probability theory, computer science, and statistics[4].

The BN model uses probability that is called Bayes' theorem or rule which is used to calculate the posterior probability which is a conditional probability of future uncertain event based on relevant evidence relating to it historically.

Bayesian Network Model uses Naive Bayes which is a type of supervised-learning module that contains examples of the input-target mapping that the model tries to learn. Such models make predictions about new data based on the examination of previous data.

Bayes' Theorem states that the probability of a particular predicted event, given the evidence in this instance, is computed from three other numbers: the probability of that prediction in similar situations in general, ignoring the specific (the so called prior probability) multiplied with the probability of seeing the evidence we have here, given that the particular prediction is correct divided by the probability of that prediction in general.

Bayesian prediction has been used in predicting music sales and has been proved to be meaningful approach in this regard. Bayesian prediction has been used in predicting music sales and has been proved to be meaningful approach in this regard.

A hierarchical Bayesian model developed based on a logistic diffusion process that allowed for the generalization of various adoption patterns out of discrete data and was applied in a situation where the eventual number of adopters was unknown. It was based on prelaunch data such as the success of previous records and updated sequentially when the first sales data of the participant record were available [5].

Naïve Bayesian learning has been adversely used to study adoption behaviors in social network by predicting individuals' adoption probabilities from observed adoption data. They identified key factors and used real-world data from two large-scale social networks to produce empirical evidence that reveals greater predictive power.[6]

2.1 Advantages of Bayesian classification

The Bayesian model has the following advantages as compared to the other models that are used for prediction.

a) Consistent, theoretically solid mechanism for processing uncertain information

Probability theory provides a consistent calculus for uncertain inference, meaning that the output of the system is always unambiguous. Given the input, all the alternative mechanisms for computing the output with the help of a Bayesian network model produce exactly the same answer.

b) Smoothness properties

Bayesian network models have been found to be very robust in the sense that small alterations in the model do not affect the performance of the system dramatically. This means that maintaining and updating existing models is easy since the functioning of the system changes smoothly as the model is being modified. For sales and marketing systems this is a crucial characteristics, as these systems need to be able to follow market changes rapidly without complex and time consuming re-modeling.

c) Flexible applicability

Bayesian networks model the problem domain as a whole by constructing a joint probability distribution over different combinations of the domain variables. This means that the same Bayesian network model can be used for solving both discriminative tasks (classification) and regression problems (configuration problems and prediction). Besides predictive purposes, Bayesian networks can also be used for explorative data mining tasks by examining the conditional distributions, dependencies and correlations found by the modeling process.

d) A theoretical framework for handling expert knowledge

In Bayesian modeling, expert domain knowledge can be coded as prior distributions, prior meaning that the probability distributions are defined before and independently of processing any possible sample data. This allows for combining expert knowledge with statistical data in a very practical way. Using suitable prior distributions, the priors can be given a semantically clear explanation in terms of the data (expert knowledge can be interpreted as an unseen data-set of the same form as the training data). This means that the experts will also be able to give an estimate of the weight or importance of their prior knowledge, compared to the training data available.

e) A clear semantic interpretation of the model parameters

Unlike neural network models, which usually appear to the user as a ``black box", all the parameters in Bayesian networks have an understandable semantic interpretation. It is for this reason that Bayesian networks can be constructed directly by using domain expert knowledge, without a time-consuming learning process. On the other hand, if machine learning techniques are used (with or without expert knowledge) for constructing Bayesian network models from sample data, the resulting model can be analyzed and explained in terms that are understandable to domain experts.

f) **Different variable types**

Probabilistic models can handle several different type variables at the same time, whereas many alternative model technologies have been designed for some single specific type of variables (continuous, discrete etc.). For these alternatives, working with several variable types requires some kind of transformation operations, which in some cases may be the cause for unexpected results. From the probabilistic point of view, all the basic entities are distributions, which mean that all the different variable types fall elegantly in the same unifying framework.

g) A theoretical framework for handling missing data

In the Bayesian network model, missing data is marginalized out by integrating over all the possibilities of the missing values. Although the advantages of probabilistic modeling have been largely recognized and accepted, the probabilistic approach has often been neglected in the past as the theoretically correct, but computationally infeasible methodology. Perhaps the most common argument against using probabilistic models has been that the number of parameters needed for defining the models is too high. Nevertheless, the theoretical framework of Bayesian network modeling suggests that it is possible to construct quite successful probabilistic models using only a moderate number of parameters. In addition, Bayesian networks appear to be rather insensitive to the accuracy of the parameters, so determining good parameter values is in many application areas quite feasible.

2.2 Application Areas of Bayesian Networks

Bayesian networks have been applied in various fields such as data mining, troubleshooting, bioinformatics/computational biology and medical diagnosis [7]in cost minimizing troubleshooting strategies, in another scenario to recognize words representing the names of Tunisian cities [8], a bayesian network model was develop and the work was based on a dynamic hierarchical Bayesian network. The aim was to find the best model of Arabic handwriting to reduce the complexity of the recognition process by permitting the partial recognition.

To test the impact of the choice of cut-off points and sampling procedures, three bankruptcy prediction models were assessed Bayesian, Hazard and Mixed Logit. However, when tests were conducted with business-cycle samples, the Bayesian model has the best performance and much better predictive power in recent business cycles[9].This study extended on recent research comparing the performance of bankruptcy prediction models by identifying under what conditions a model performs better. It also allays a range of user groups, including auditors, shareholders, employees, suppliers, rating agencies, and creditors' concerns with respect to assessing failure risk.

Prior research has shown that BNs perform well as a classification and prediction tool in different domains [8] Unlike most regression techniques, BNs do not have any requirements on the underlying distributions of variables. BNs can easily model complex relationships among variables including partial mediators and "interaction effects". BNs do not require complete information for observations, Observations that have some missing variables can still be used to train or test BN models,BNs are dynamic and interactive and they can easily be updated with new information as it is learned. Subjective human knowledge can easily be incorporated into models as compared to other machine learning techniques, such as neural networks, BN models are more transparent and intuitive because relationships among variables are explicitly represented by the direct acyclic[10].

Predicting insider threats from employs in government and corporate organizations who are malicious and steal confidential information destroy systems or even kill co-workers Bayesian Network model proved to be useful[11]. The model has also been extensively used in medical research in a case of survival prediction and treatment selection in lung cancer care characterized by high levels of uncertainty. Bayesian Networks (BNs), which naturally reason with uncertain domain knowledge, was applied to aid lung cancer experts by providing personalized survival estimates and treatment selection recommendations[12].

Bayesian networks are particularly interesting because they are able to encode data from experts and patient history. They also have the benefit that the results are easily explainable as compared to other methods[13].

3. SOCIAL MEDIA

The growth in size and popularity of social media sites like Facebook and Twitter has enabled researchers to use it as another source of data for prediction, [14]examined tweets from the site twitter and made predictions about the financial success of various movies. Using only tweets that preceded the release of a movie, they found a strong correlation between the amounts of attention a movie is given and its future financial success.

The examination of the predictive ability of Internet chatter from user-generated content on blogs and forums and its impact on music album sales, [15] the research concluded that online chatter is predicted album sales during the first two weeks following the album release and the week preceding the album release. In their study [16]took a similar approach with movies and found correlation between references to movies in weblog posts both before and after their release and the movies' financial success.

Stock market investors have increasingly turned to Twitter as an investment tool. Twitter's appeal as an investment tool lies in the user's ability to relay company information, investment ideas and market sentiment in a short, concise manner.

1.5 million Messages focused on 45 companies that were posted on Yahoo! Finance and RagingBull.com, was concluded that stock market messages help to predict market volatility and that while economically small; stock market message boards do affect stock market returns in a statistically significant manner.[17]

A formal sentiment analysis tool was developed and applied to stock market message boards and found no significant correlation between sentiment and stock price movements. They acknowledged this in their research and attributed it to the large amount of noise in stock market message boards as well as the lack of market power that many investors participating in online message boards have[18].

Previous research has primarily used online message boards as a mean to aggregate investor sentiment. Recent research has taken advantage of the emergence of social media and applied this towards financial markets.

In Signaling six different mood types that would reflect the mood of an individual (Calm, Alert, Sure, Vital, Kind and Happy), [19] aggregated "tweets" from Twitter as a whole, choosing not to focus stock market specific "tweets," and examined the "mood" states of Twitter users and corresponding stock market movements. The conclusion showed the collective "mood" of Twitter users successfully predicted the upward and downward movement of the stock market.

3.1 Using Bayesian Classifier to classify Sentiments from Social media

The following sentiments were posted by buyers expressing their opinions about specific phone brands that they had purchased. The buyers' opinions show satisfaction or dissatisfaction that would directly influence the judgment of a prospect buyer of the same specific brand.

The brands that were sampled have been assigned the following serial numbers(S/No) to conceal the identity of the brand names.

4. Tables and Equations

4.1 Tables

Table 1: Brand parameters table

Brand	Features (Resolution, OS, Screen,)	Hardware (Battery Life, RAM, Processor)	Probability of high preference
S/No1	Camera Clarity, OS fast, no apps, lack of top level apps, screen clear, great clarity, its fast, built ability of features, resolution worst	Battery life long, processor fast, it shuts processor issue frustrating, processor overheats Extra memory slot	Yes
S/No2	Camera good, awesome features, updates of apps, good resolution, browser ok.	Processor excellent, big memory, processor fast, plastic, slow to boot	Yes
S/No3	Resolution poor, can't achieve camera clarity same level phone, awful features, browser ok	Long battery life processor hangs and restarts, processes slowly, Have nothing special	No
S/No4	camera pixels wow, other features not bad, camera pixels wow, pics are blurred. Not good for heavy games	Takes ages to load RAM, Processor slow, and battery keeps power, cheap and average smart phone,	Yes
S/No5	Camera good, resolution good, can't play movies, camera crystal clear	Processor slow, RAM too small, Battery life short, big screen,	No

Table 2: likes and dislikes table

Brand	Likes	Dislikes
S/No1	Camera clarity, windows OS fast, long battery life, processor fast, excellent camera, great size , extra memory	processor overheats, no apps, no top level apps
S/No 2	Processor Excellent, big memory, Camera good	Android slow to boot,
S/No 3	Long battery	camera clarity disappointment, life Resolution poor, processor hangs and restarts
S/No 4	camera pixels wow, battery keeps power, cheap and average smart phone,	Takes ages to load RAM, Processor slow, pics blurred,
S/No 5	Camera good, big screen, resolution good, camera crystal clear	Processor slow, RAM too small, Battery life short,,, can't play movies.

The sentiments have been classified in the table below using features, hardware and operating system

Features (Good, Moderate, Awful)

Hardware (Excellent, Moderate, Bad)

Operating system (Windows, Android)

Table: 3 Classification table

Brand	Twit	Features	Hardware	Operating system	class
S/No 1	Twit1	Good	Bad	Android	Yes
	Twit7	Good	Excellent	Android	Yes
	Twit 8	Good	Good	Android	Yes
	Twit 28	Moderate	Good	Android	Yes
S/No 2	Twit 2	Good	Good	Windows	Yes
	Twit 9	Good	Good	Windows	Yes
	Twit 13	Good	Bad	Windows	Yes
	Twit 16	Good	Good	Windows	Yes
	Twit 17	Awful	Bad	Windows	No
S/No 3	Twit 18	Awful	Bad	Android	No
	Twit 23	Moderate	Moderate	Android	Yes
	Twit 24	Moderate	Bad	Android	No
	Twit 25	Good	Good	Android	Yes
S/No 4	Twit 19	Moderate	Bad	Android	No
	Twit 20	Good	Moderate	Android	Yes
	Twit 21	Awful	Bad	Android	No
	Twit 22	Awful	Good	Android	Yes
S/No 5	Twit 3	Good	Good	Android	Yes
	Twit 6	Moderate	Bad	Android	No
	Twit 11	Awful	Bad	Android	No
	Twit 15	Good	Bad	Android	Yes

4.2 Equations

The Naïve Bayes algorithm uses the mathematics of Bayes' Theorem to make its predictions. The theorem is denoted by

$$P(A/B) = \frac{P(B/A).\ P(A)}{P(B)} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

Computation of the classified data

$$v_{NB} = \arg\max_{v_j \epsilon V} \; P(v_j) \prod_{i=1}^{n} P\,(a_i/_{V_j})$$

$$v_{NB} = \arg \max_{v_j \epsilon \{yes,no\}} \; P(v_j)P(features = Awful \backslash v_{j)}.\ P(hardware = moderate \backslash v_{j)}.\ P(operating\ system = android \backslash v_{j)}$$

$$(Probability\ of\ high\ preference = yes) = P(yes) = {}^{14}/_{21} = 0.67$$

$$(Probability\ of\ high\ preference = no) = P(no) = {}^{7}/_{21} = 0.33$$

$$P(Hardware = moderate \backslash\ high\ preference = yes) = p(moderate \backslash yes) = 2/14 = 0.143$$

$$P(Hardware = moderate \backslash\ high\ preference = no) = p(moderate \backslash no) = {}^{3}/_{7} = 0.429$$

P (moderate\no = n_c/n, where n=7 is the total number of training examples for which probability of high =no preference and n_c =3 is the number in which moderate = no.

We can also compute the probability of the following

Bad\yes= Android\yes= Excellent\yes=

Bad\no= Android\no= Good\no=

$$P(yes).P(bad \backslash yes).P(Android|yes).P(Excellent \backslash yes) = \frac{14}{21} * \frac{3}{14} * \frac{10}{14} * \frac{1}{14} = 0.007$$

$$= P(no).P(bad \backslash no).P(Android|no).P(Excellent \backslash no) = \frac{7}{21} * \frac{7}{7} * \frac{6}{7} * \frac{1}{7} = 0.04$$

5. Conclusions

The Internet and other digital networks are the driving forces behind a dramatic change in the way business dealings are conducted. Increasingly, social networks are being used to electronically design, market, buy, sell, and deliver products and services worldwide. This Internet Economy has seen tremendous growth and with latest fastest internet connectivity technologies has made internet service to be available to almost all consumers; this has presented all marketers to trade via e-commerce.

The expansion of the Internet in the past decade has given researchers new avenues to explore the art of prediction. First, online forums and blogs allowed individuals to share thoughts; opinions and information with one another on any imaginable subject as a result social media platforms were formed. We reviewed some research work on prediction models in order to understand the current research on Bayesian prediction: where we looked at other predictive models that exist in literature.

We sampled opinions from social media and we classified them using the attributes that influenced the buyers of certain brands, then we computed the probability of high/low preferences of a brand based on these attributes.

In this paper we have used a Bayesian classifier as a predictive model using phone attributes sentiments from social media for predicting sales. However there may be other features or attributes that may influence the preference of a brand. Future research works may look at the application of Bayesian Network Model in predicting marketing trends of those products whose buyers are fond of using social media to express their affection, experiences, satisfaction or dissatisfaction towards recently bought products.

References

[1] V. Dhar and E. A. Chang, "Does chatter matter? The impact of user-generated content on music sales," *Journal of Interactive Marketing*, vol. 23, pp. 300-307, 2009.

[2] G. Mishne and N. S. Glance, "Predicting Movie Sales from Blogger Sentiment," in *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, 2006, pp. 155-158.

[3] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, 2010, pp. 492-499.

[4] I. Ben-•Gal, "Bayesian networks," *Encyclopedia of statistics in quality and reliability*, 2007.

[5] K.E. Lee, Gene Selection a Bayesian Variable Selection Approch . Bioinformatics, 19,90-97, 2003.

[6] X. Fang, P. J.-H. Hu, Z. Li, and W. Tsai, "Predicting adoption probabilities in social networks," *Information Systems Research*, vol. 24, pp. 128-145, 2013.

[7] T. Gustavsson, Troubleshooting using Cost effective Algorithms and Bayesian Networks. Masters degree Project Stockholm, Sweden, XR-EE-RT 2007:2. 2006.

[8] K. Jayech, N. Trimech, M. A. Mahjoub, and N. E. Ben Amara, "Dynamic hierarchical Bayesian network for Arabic handwritten word recognition," in *Information and Communication Technology and Accessibility (ICTA), 2013 Fourth International Conference on*, 2014, pp. 1-6.

[9] L. Sun, and P.P. Shenoy,A. Using Bayesian Network for Bankruptcy Prediction , European Journal of Operational Research 180 vol 2, pp. 738-753. 2007.

[10] S. Trabelsi, R. He, L. He, and M. Kusy, "A comparison of Bayesian, Hazard, and Mixed Logit model of bankruptcy prediction," *Computational Management Science*, pp. 1-17, 2013.

[11] R.D Anderson, R.D. Mackoy, V.B. Thompson and G. Harrel A bayesian Network Estimation of the Service-Profit Chain for transport service Satisfaction Decision Sciences, 2004. 35 vol 4 pp. 665-690

[12] E. T. Axelrad, P. J. Sticha, O. Brdiczka, and J. Shen, "A Bayesian network model for predicting insider threats," in *Security and Privacy Workshops (SPW), 2013 IEEE*, 2013, pp. 82-89.

[13] M. B. Sesen, A. E. Nicholson, R. Banares-Alcantara, T. Kadir, and M. Brady, "Bayesian networks for clinical decision support in lung cancer care," *PloS one*, vol. 8, p. e82349, 2013.

[14] A. Madsen, "Bayesian Networks for Disease Diagnosis," *Unpublished Manuscript. B. Thomas Golisano College of Computing and Information Sciences. Rochester, NY*, 2010.

[15] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, 2010, pp. 492-499.

[16] W. Antweiler and M. Z. Frank, "Is all that talk just noise? The information content of internet stock message boards," *The Journal of Finance*, vol. 59, pp. 1259-1294, 2004.

[17] S. R. Das and M. Y. Chen, "Yahoo! for Amazon: Sentiment extraction from small talk on the web," *Management Science*, vol. 53, pp. 1375-1388, 2007.

[18] J. Bollen, H. Mao, and A. Pepe, "Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena," in *ICWSM*, 2011.

First Author Everlyne Wamukekhe is a holder of BSc. Information Technology (2011), Ongoing MSc. Software Engineering Jomo Kenyatta University of Agriculture and Technology.
Worked at Kenya Institute of Education and curriculum development (2001) and currently working at Jomo Kenyatta University of Agriculture and Technology.
as an ICT officer and a CISCO Instructor department of Computing

Second Author Prof. Waweru Mwangi is an associate professor and a Senior Lecturer in the Department of Computing, Jomo Kenyatta University of Agriculture and technology. He is an Associate Professor of Information Systems Engineering.
PhD in Information Systems Engineering, Hokkaido University (Japan) 2004, MSc Operation s Research and Cybernetics, Shanghai University (China) 1995. Bed Mathematics Kenyatta University (Kenya) 1989.
Director Institute of Computer Science and Information Technology in JKUAT 2004 to 2013.
Research interests: Information architecture, System performance optimization, Simulation and modeling , Intelligent agent modeling.
Publications; Published more than 20 papers in international and local journals related to the field of ICT.

Third Author Dr. Michael W. Kimwele is a Lecturer in the Department of Computing, Jomo Kenyatta University of Agriculture and technology (JKUAT). He holds a BSc. Mathematics and Computer Science-First Class Honours from JKUAT (2002), a Masters in Information Technology Management from University of Sunderland-UK (2006) and a Doctorate in Information Technology from JKUAT (2012). At present, he is the Associate Chairman, Department of Information Technology; JKUAT-Westlands Campus.Dr.Kimwele has authored a commendable number of research papers in international/national conference/journals and also supervises postgraduate students in Computer Science/Information Technology. His research interests include Information systems management, Information Technology Security, Electronic Commerce, Mobile Computing, Social implications of computer applications, Human Computer Interaction, and Computer Ethics.