**LITERATURE SURVEY**

Title: Fundamentals of sentiment analysis and its applications.

Abstract: The problem of identifying people’s opinions expressed in written language is a relatively new and very active field of research. Having access to huge amount of data due to the ubiquity of Internet, has enabled researchers in different fields—such as natural language processing, machine learning and data mining , text mining , management and marketing and even psychology—to conduct research in order to discover people’s opinions and sentiments from the publicly available data sources. Sentiment analysis and opinion mining are typically done at various level of abstraction: document, sentence and aspect. Recently researchers are also investigating concept-level sentiment analysis , which is a form of aspect-level sentiment analysis in which aspects can be multi terms. Also recently research has started addressing sentiment analysis and opinion mining by using, modifying and extending topic modeling techniques. Topic models are probabilistic techniques for discovering the main themes existing in a collection of unstructured documents. In this book chapter we aim at addressing recent approaches to sentiment analysis, and explain this in the context of wider use. We start the chapter with a brief contextual introduction to the problem of sentiment analysis and opinion mining and extend our introduction with some of its applications in different domains. The main challenges in sentiment analysis and opinion mining are discussed, and different existing approaches to address these challenges are explained. Recent directions with respect to applying sentiment analysis and opinion mining are discussed. We will review these studies towards the end of this chapter, and conclude the chapter with new opportunities for research.

Title: Importance of Customer Satisfaction

Abstract: Customer satisfaction is important because it illustrates whether your customer base likes what you’re doing. [Research shows](https://www.zendesk.com/blog/report-cx-maturity-north-america/) that high satisfaction leads to greater customer retention, higher lifetime value, and a stronger brand reputation.

Low customer satisfaction scores are important, too. They can reveal customer pain points and provide data-backed insights on how to improve your product, service, and overall customer experience.

Title: Sentiment Analysis and Opinion Mining

Abstract: Opinions are central to almost all human activities and are key influencers of our behaviors. Our beliefs and perceptions of reality, and the choices we make, are, to a considerable degree, conditioned upon how others see and evaluate the world. For this reason, when we need to make a decision we often seek out the opinions of others. This is not only true for individuals but also true for organizations

Title: Random Forest.

Abstract: Random forest is a commonly-used machine learning algorithm trademarked by Leo Breiman and Adele Cutler, which combines the output of multiple decision trees to reach a single result. Its ease of use and flexibility have fueled its adoption, as it handles both classification and regression problems.

Title: Thai sentiment analysis for consumer’s review in multiple dimensions using sentiment compensation technique (SenSecomp).

Abstract: Trustworthiness of an e-vendor in the e-marketplace can be determined in multiple dimensions: product, price, and shipping. An e-vendor who has high trust level in more dimensions is more likely to have the competitive advantage than others. A consumer's review is analyzed to find i ts p olarity in different dimensions. Positive sentiment in consumers' reviews helps increase the trustworthiness of e-vendors which in turn influences c onsumer's p urchase i ntention. I n t his p aper, we propose the method to automatically analyze Thai sentiment of consumer's review in product, price, and shipping dimensions by using multi-dimensional lexicon and sentiment compensation technique. A consumer's review in Thai language is tokenized using the longest matching algorithm. Then, it is analyzed to find its sentiment. Sentiment compensation technique is used to automatically compensate the sentiment to a dimension where consumer's review mentions the sentiment without a dimension. The results show that our proposed method outperform sentiment to dimension (S2D) and dimension to sentiment (D2S) methods with the overall accuracy 93.60%.

Title: Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine.

Abstract: AbstrakAnalisis sentimen dalam penelitian ini merupakan proses klasifikasi dokumen tekstual ke dalam dua kelas, yaitu kelas sentimen positif dan negatif. Data opini diperoleh dari jejaring sosial Twitter berdasarkan query dalam Bahasa Indonesia. Penelitian ini bertujuan untuk menentukan sentimen publik terhadap objek tertentu yang disampaikan di Twitter dalam bahasa Indonesia, sehingga membantu usaha untuk melakukan riset pasar atas opini publik. Data yang sudah terkumpul dilakukan proses preprocessing dan POS tagger untuk menghasilkan model klasifikasi melalui proses pelatihan. Teknik pengumpulan kata yang memiliki sentimen dilakukan dengan pendekatan berdasarkan kamus, yang dihasilkan dalam penelitian ini berjumlah 18.069 kata. Algoritma Maximum Entropy digunakan untuk POS tagger dan algoritma yang digunakan untuk membangun model klasifikasi atas data pelatihan dalam penelitian ini adalah Support Vector Machine. Fitur yang digunakan adalah unigram dengan fitur pembobotan TFIDF. Implementasi klasifikasi diperoleh akurasi 86,81 % pada pengujian 7 fold cross validation untuk tipe kernel Sigmoid. Pelabelan kelas secara manual dengan POS tagger menghasilkan akurasi 81,67%. Kata kunci—analisis sentimen, klasifikasi, maximum entropy POS tagger, support vector machine, twitter. AbstractSentiment analysis in this research classified textual documents into two classes, positive and negative sentiment. Opinion data obtained a query from social networking site Twitter of Indonesian tweet. This research uses Indonesian tweets. This study aims to determine public sentiment toward a particular object presented in Twitter businesses conduct market. Collected data then prepocessed to help POS tagged to generate classification models through the training process. Sentiment word collection has done the dictionary based approach, which is generated in this study consists 18.069 words. Maximum Entropy algorithm is used for POS tagger and the algorithms used to build the classification model on the training data is Support Vector Machine. The unigram features used are the features of TFIDF weighting.Classification implementation 86,81 % accuration at examination of 7 validation cross fold for the type of kernel of Sigmoid. Class labeling manually with POS tagger yield accuration 81,67 %. Keywords—sentiment analysis, classification, maximum entropy POS tagger, support vector machine, twitter.

Title: Predicting the Timeliness of Student Graduation Using Decision Tree C4

Abstract: The purpose of this study is to predict whether a student will graduate on time or not, using the data collected from the database of Academic Administarion Bureau of Universitas Advent Indonesia (UNAI) for the academic period of 2009-2013. There are 9 (nine) attributes that were used to predict the timeliness factors, and the method used for predicition is the C4.5 decision tree. This study also used SMOTE (Synthetic Minority Oversampling Technique) in the WEKA application to balanced data in minor class. The result shows that the attribute GPA, was the highest root that gave the highest influence. Another attributes that also has influence in predicting the student's graduation timelines are: repeating courses, study leave, and gender, while the attribute religion, is the attributes that are less influential. The accuracy on cross validation 10 Folds with SMOTE and without SMOTE the results shows that: with SMOTE 83,055% and without SMOTE 82,644%. Split Test 70:30 with SMOTE and without SMOTE, the results are 82,026% and 84,015%. The use of SMOTE increases the value of precision and recall. The value of precision for not ontime student, with SMOTE 82.5% and without SMOTE 76.4%, the value of recall with SMOTE 80.6% and without SMOTE 61.9%.

Title: PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE.

Abstract: The quality of lectures can be determined by some feedbacks from students. From the feedbacks, we can give appreciations for those lectures who get good feedback from students, and evaluations for those who get bad feedback. The problem is classifying large size of feedbacks manually isn’t effective and took a lot of time. Therefore, we need a system that can classify feedbacks automatically. These feedbacks will be classified into positive, negative, and neutral, usually called as sentiment analysis. Sentiment analysis implementation can be done by several methods, one of them that has a good accuracy is Support Vector Machine (SVM). SVM performance in this research is measured with the level of accuracy. The number of accuracy indicate the success level of system. The conclusion of this research is factors that affects the accuracy. The factors are the range of each classes and number of unique words in the training document.

Title: Research on Aero-Engine Maintenance Level Decision Based on Improved Artificial Fish-Swarm Optimization Random Forest Algorithm. Proceedings

Abstract: In this paper, an aero-engine exhaust gas temperature (EGT) prediction model based on LightGBM optimized by the chaotic rate bat algorithm (CRBA) is proposed to monitor aero-engine performance effectively. By introducing chaotic rate, the convergence speed and precision of bat algorithm are im-proved, which CRBA is obtained. LightGBM is optimized by CRBA and it is used to predict EGT. Taking a type of aero-engine for example, some relevant performance parameters from the flight data measured by airborne sensors were selected as input variables and EGT as output variables. The data set is divided into training and test sets, and the CRBA-LightGBM model is trained and tested, and compared with ensemble algorithms such as RF, XGBoost, GBDT, LightGBM and BA-LightGBM. The results show that the mean absolute error (MAE) of this method in the prediction of EGT (after normalization) is 0.0065, the mean absolute percentage error (MAPE) is 0.77% and goodness of fit R2 has reached to 0.9469. The prediction effect of CRBA-LightGBM is better than other comparison algorithms and it is suitable for aero-engine condition monitoring.

Title:

Abstract: