

# Fake Review Detection Using Machine Learning Techniques

Shilpa yadav<sup>1</sup>, Dr.Gulbakshee Dharmela<sup>2</sup>, Khushali Mistry<sup>3</sup>  
Post Graduate Student, Parul University, vadodara, Gujrat

<sup>2</sup>Professor, Parul University, vadodara, Gujrat, <sup>3</sup>Professor, Parul University, vadodara, Gujrat

**Abstract:** Online reviews play a very important role in today's e-commerce for decision-making. Large part of the population i.e. customers read reviews of products or stores before making the decision of what or from where to buy and whether to buy or not. As writing fake/fraudulent reviews comes with monetary gain, there has been a huge increase in deceptive opinion spam on online review websites. Basically fake review or fraudulent review or opinion spam is an untruthful review. Positive reviews of a target object may attract more customers and increase sales; negative review of a target object may lead to lesser demand and decrease in sales. These fake/fraudulent reviews are deliberately written to trick potential customers in order to promote/hype them or defame their reputations. Our work is aimed at identifying whether a review is fake or truthful one.

**Index Terms** – Naïve Bayes (NB), Support Vector Machine (SVM), K-Nearest Neighbors (KNN-IBK), KStar (K\*) and Decision Tree(DT)

## I. INTRODUCTION

Reviews are statements which express suggestion, opinion or experience of someone about any market product. On the online e-commerce websites, users place their reviews on product form to give suggestion or share experience with product providers / sellers / producers and new purchasers. The provided user experience can help any business to grow for improvement by analyzing the suggestions. Polarity of reviews causes certain financial gain or loss to any product provider.

On other side, reviews influence new purchasers while taking decision of purchasing any particular product. It can be concluded that effects of reviews target both business and users in different ways. Keeping this point of view, many firms / product providers hire agents to forge fake opinions for growing their business and market reputation. As a result, users take wrong product selection decision. The pattern of web based shopping is developing day by day. Online e-commerce websites opened channel for selling or purchasing products.

E-commerce sites facilitates users to purchase product (e.g. motor bike, headphones, laptop, etc.) or avail any service (i.e. hotel reservation, airline ticket booking, etc.). Users often give suggestion/opinion/review/comment on e-commerce sites to share their experience after using any product or availing service. BCI helps by using the brain thoughts as input signals for applications such as cursor control, robotic arms, wheelchairs, and other devices.[2]

Opinion spamming is an immoral activity of posting fake reviews. The goal of opinion spamming is to misguide the review readers. Users involved in spamming activity are called “spammers”. The task of a spammer is to build fake reputation (either good or bad) of a business by placing fake reviews.

### 1.1.1 Importance of User Reviews

Online purchasers on e-commerce sites are increasing day by day. Online purchasers often post reviews/opinions about certain product they have used. In other words, opinions are content created by users on e-commerce websites to express experience of users about any service or product.

Importance of user reviews can be viewed from user and business perspective. From user perspective, these reviews can influence new customers/users for purchasing decision of certain product in a good or bad way. Decision of new purchasers is influenced by reviews of users. Good or bad features in accordance with user experience are described in reviews which help other users for taking the decision of purchasing the product. For purchasing online, user often visit e-commerce sites rich with user experience about products. So quality and number of user experience can effect user traffic on site.

### 1.1.2 What is Fake Review?

Opinion spamming is an immoral activity of posting fake reviews. The goal of opinion spamming is to misguide the review readers. Users involved in spamming activity are called “spammers”. The task of a spammer is to build fake reputation (either good or bad) of a business by placing fake reviews.

There exist some businesses who pay spammers to promote the company to attract new customers or to demote competent company of same type of business. A fake review either belong to positive or negative polarity. Review containing praising statement about the product fall in “positive polarity”. And review containing loathing statements about the product fall in “negative polarity”.

Increasing need for identifying fake reviews has captured the attention of researchers for solving the problem. Fake reviews not only mislead new customer for taking product purchasing decision but also affects business of good quality product. And due to false and misleading reviews on particular e-commerce site, users will avoid to visit that particular e-commerce site. It is concluded that identifying fake reviews will tackle three loses at one time.

### 1.1.3 Contextual and Behavioral Features

It is reported by researchers that the task of identifying untruthful reviews is more challenging task than identifying brand reviews and non-reviews (D. Zhang, Zhou, Kehoe, & Kilic, 2016). Commonly, two types of features are used to identifying fake reviews: Contextual and Behavioral features.

### 1.1.4 Background Knowledge

Fake review detection task is one of the challenging classification task in the field of knowledge discovery. Multiple angles of capturing deception in reviews data have been focused by researchers for a decade. Focus of our research work is to investigate the techniques and classification model to identify individual fake reviews by analyzing different perspective of review data.

### 1.1.5 Data Mining Techniques

Generally, DM tasks can be divided into two groups: Descriptive mining and Predictive mining (U. Fayyad et al., 1996; Heydari et al., 2015; Crawford et al., 2015). Descriptive mining involves describing the general characteristics of the information in the database i.e. clustering and association rules whereas predictive mining involves forecasting values on the basis of available current data i.e. regression, classification and analysis of outlier (Berry & Linoff, 1997; J. Han, Pei, & Kamber, 2011). We define some of general techniques of data mining. The section 2 of this paper will give the literature survey, section 3 gives the proposed work and section 4 gives the conclusion.

## II. LITERATURE SURVEY

The task of fake review detection has been studied since 2007, with the analysis of review spamming [1]. In this work, the authors analyzed the case of Amazon, concluding that manually labeling fake reviews may result challenging, as fake reviewers could carefully craft their reviews in order to make them more reliable for other users. Consequently, they proposed the use of duplicates or nearly-duplicates as spam in order to develop a model that detects fake reviews [1]. Research on distributional footprints has also been carried out, showing a connection between distribution anomalies and deceptive reviews from Amazon products and TripAdvisor hotels [2].

Fake review detection is a specific application of the general problem of deception detection, where both verbal and nonverbal clues can be used [3]. Fake review detection research has mainly exploited textual and behavioral features, while other approaches have taken into account social or temporal aspects. Textual features have been proposed in several papers. Ott et al. [4] employed psycholinguistic features based on LIWC [5] combined with standard word and Part of Speech (POS) n-gram features. Mukherjee et al. [6] extend that work including also style and POS based features, such as deep syntax and POS sequence patterns.

Behavioral features refer to nonverbal characteristics of review activity, such as the number of reviews or the time and device where the review was posted. They were used in order to improve the classification model resulting in encouraging results. Liu et al. [31] introduced behavioral features on Amazon reviews, distinguishing among review features (e.g. number of feedbacks, position of the review, textual features, rating features, etc.), product features (e.g. price, sales rank) and reviewer features (e.g. average rating, ratio of the number of reviews that the reviewer wrote which were the first reviews, etc.).

In another work, Zhang et al. [58] explore the effect of both textual and behavioral features in the restaurant and hotel domain, showing that non-textual features result more relevant for the task of fake review detection. Apart from using textual and behavioral features, other methodologies were followed for the fake review detection task. Wang et al. [55] proposed a review graph with the aim of capturing relationships between reviewers, reviews and stores reviewed by the reviewers. Making use of this graph, an iterative model was used to identify suspicious reviewers. Following also a graph model, network effects were analyzed by Akoglu et al. [1], following two steps: user and review scoring for fraud detection and grouping for visualization.

Another methodological approach focuses on temporal aspects, and concerns the burstiness of reviews and their impact on businesses. Bursts of reviews can be either due to sudden popularity of products or spam attacks [17], which were also analyzed in [38] along with other behavioral and textual features. A deeper time series approach was made by Heydari et al. [27] and Li et al. [7] propose other types of features such as review density in temporal windows, along with semantic and emotion features. Spatial and temporal features were used in a Chinese site by Li et al. [36].

Regarding classification algorithms, Support Vector Machine [54] was the most used one followed by Naive Bayes [22], Decision Tree [5], Random Forest [4] and Logistic Regression [11]. Apart from supervised learning, other approaches have been followed, since collecting data for experiments is a hard task. In [35], authors propose a prediction model based on semi-supervised learning and a set of textual and behavioral features. Additionally, Hernandez et al. [23] propose a semi-supervised technique called PU-learning.

### III. PROPOSED SYSTEM

The goal of this article is analyzing the fake review problem in the consumer electronics field, more precisely studying Yelp businesses from four of the biggest cities of the USA. No prior research has been carried out in this concrete field, being restaurants and hotels the most previously studied cases. We want to prove that fake review detection problem in online consumer electronics retailers can be solved by machine learning means and to show if the difficulty of achieving it depends on geographic location.

In order to achieve this goal, we have followed a principled approach. Based on literature review and experimentation, a feature framework for fake review detection is proposed, which includes some contributions such as the exploitation of the social perspective. This framework, so called Fake Feature Framework (F3), helps to organize and characterize features for fake review selection. F3 considers information coming from both the user (personal profile, reviewing activity, trusting information and social interactions) and review elements (review text), establishing a framework with which categorize existing research.

In order to evaluate the effectiveness of the features defined in F3, a dataset from the social Yelp in four different cities has been collected and a classification model has been developed and evaluated.

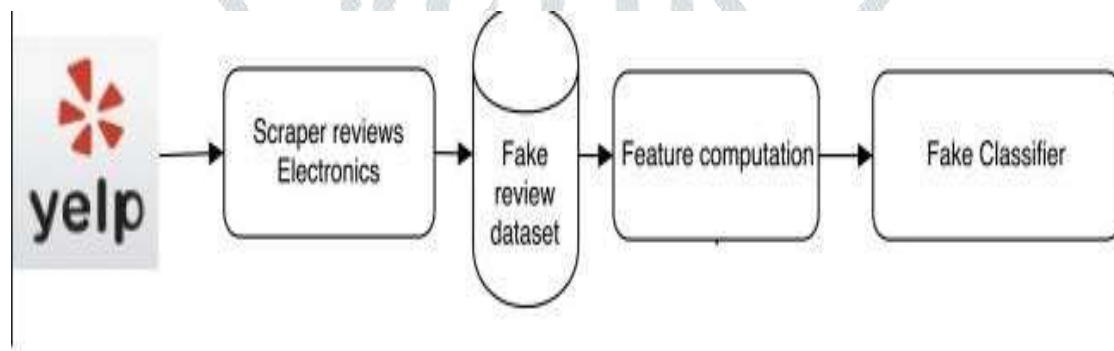
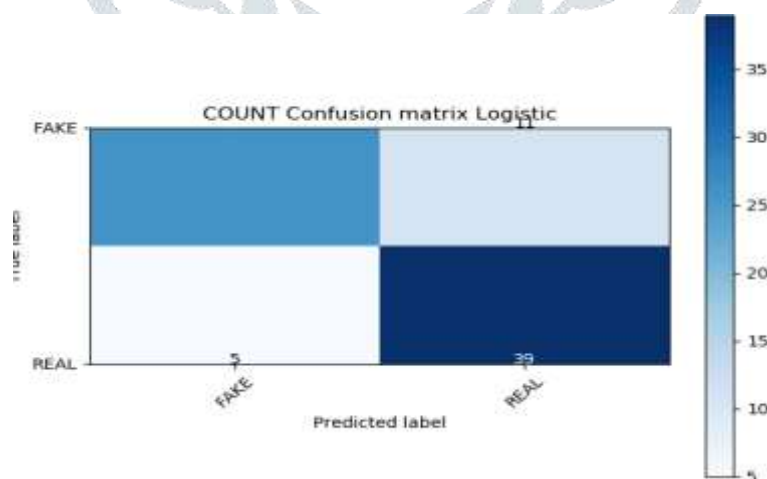


Figure : Architecture of Implementation



Data Visualization of confusion matrix

### IV. Data preprocessing

The purpose of preprocessing is to convert raw data into a form that fits machine learning. Structured and clean data allows a data scientist to get more precise results from an applied machine learning model. The technique includes data formatting, cleaning, and sampling.

## V. Dataset splitting

A dataset used for machine learning should be partitioned into three subsets — training, test, and validation sets. Training set. A data scientist uses a training set to train a model and define its optimal parameters it has to learn from data. Test set. A test set is needed for an evaluation of the trained model and its capability for generalization. The latter means a model's ability to identify patterns in new unseen data after having been trained over a training data. It's crucial to use different subsets for training and testing to avoid model overfitting, which is the incapacity for generalization we mentioned above.

## VI. Model training

After a data scientist has preprocessed the collected data and split it into train and test can proceed with a model training. This process entails “feeding” the algorithm with training data. An algorithm will process data and output a model that is able to find a target value (attribute) in new data an answer you want to get with predictive analysis. The purpose of model training is to develop a model.

## VII. Implementation Methodology

The proposed work is implemented in Python 3.6.4 with libraries scikit-learn, pandas, matplotlib and other mandatory libraries. We downloaded dataset from yelp.com. The data downloaded contains train set and test set separately with four two classes of label namely fake and real. The train dataset considered as train set and test dataset considered as test set. Machine learning algorithm is applied such as Naive bayes, SVM, logistic regression and random forest.

## VIII. Processing

In many databases of real world contain conflicting and noise data. The reason is that data is often collected from numerous and heterogeneous sources. Inconsistency in data results inaccurate outcomes in data mining process. One of the vital step is preprocessing of data before initiating process of data mining. There are various preprocessing methods (Y. Sun, Kamel, Wong, & Wang, 2007) to handle variety of data (cleansing, attribute reduction, tokenization, stop words removing, lemmatization, and stemming). Two types of preprocessing techniques are used for this research work: text and data preprocessing.

## IX. Text Preprocessing

Text preprocessing include data mining techniques used to transform unstructured text. Few text preprocessing techniques on our selected dataset are defined as follows:

**Tokenization:** Tokenization is task of splitting-up the review text into words (tokens). i.e. Review content is tokenized into tokens. For calculating RCS and capital diversity, tokenization is vital step to separate each word in review.

**Lemmatization:** The task of lemmatizer is to transform word with respect to morphological root word e.g. 'bought' lemmatized into 'buy'.

## X. Result Analysis

Experiments conducted with variation of behavioral and contextual feature sets explored importance of selected features for training fake review detection model. We compared results of different feature sets including three different term weighting schemes on Naive and RF. From initial experiments for exploring importance of “Review Deviation” with other behavioral and contextual features we analyze that by new feature improves accuracy. Whereas the finding based on our experimental results shows that by scaling dataset can improve the classification accuracy and f1- score. Literature on classifier comparison by (D. Zhang et al., 2016) also reports that RF outperformed other classifiers.

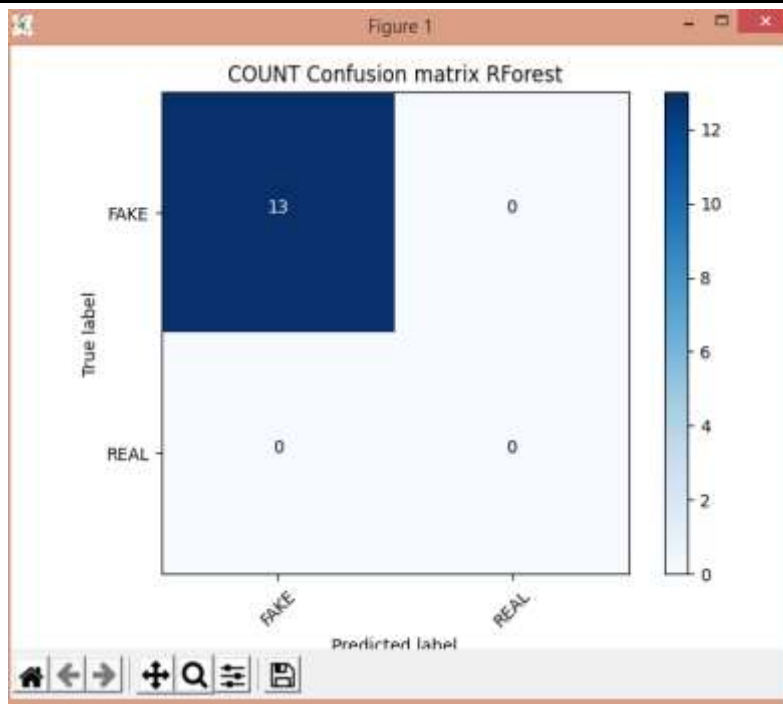


Figure 5.2.2 COUNT Confusion matrix RForest

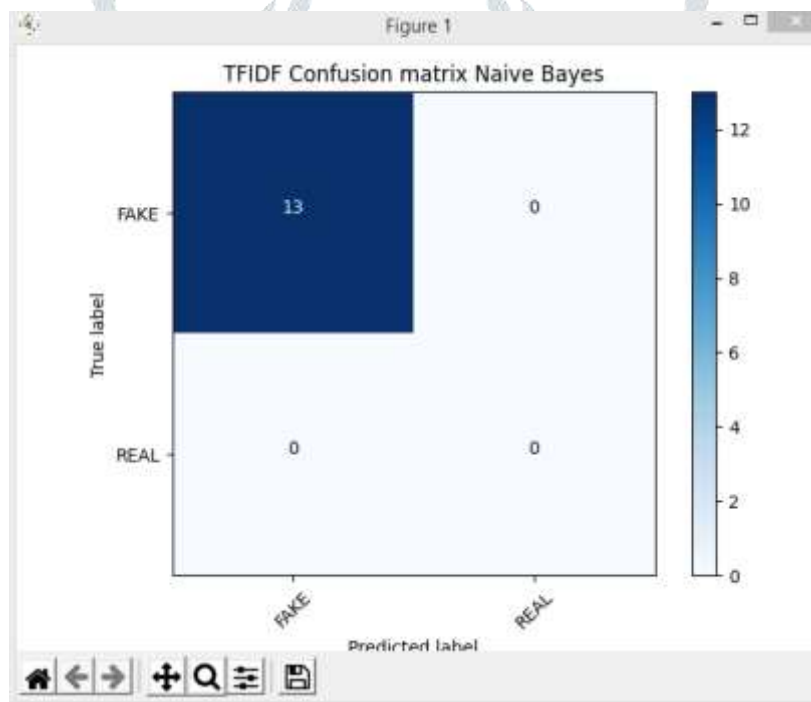


Figure 5.2.3 TFIDF Confusion matrix Naive Bayes

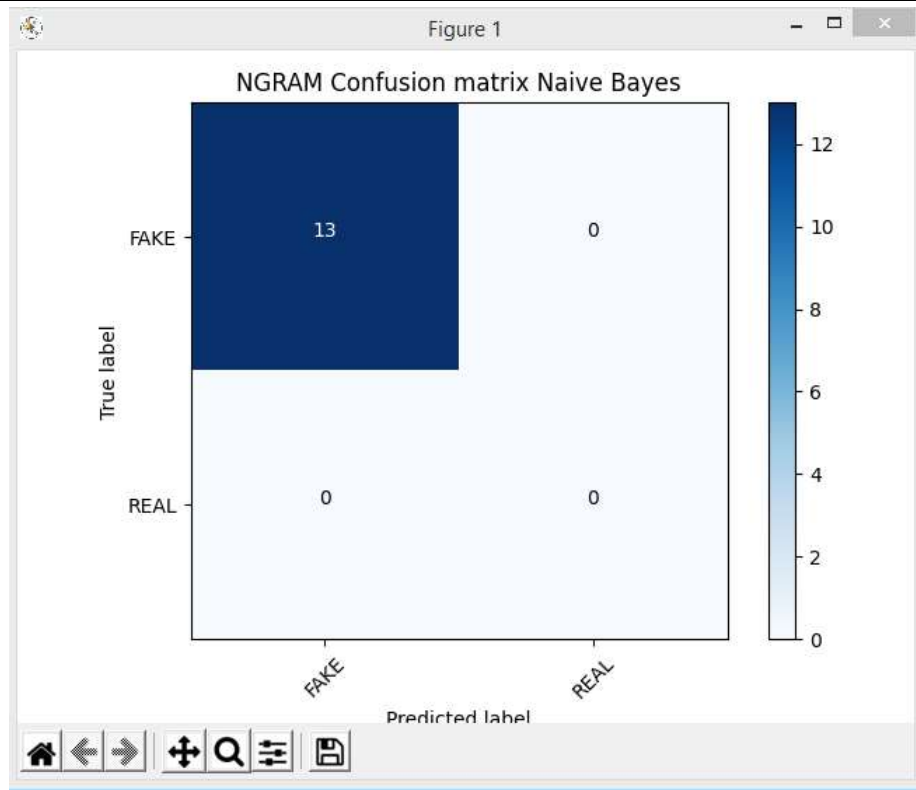


Figure 5.2.4 NGRAM Confusion matrix Naïve Bayes

## XI. CONCLUSION

We have implemented Fake review detection taken dataset by applying three feature extraction techniques namely CountVectorizer, Ngram model, TfidfVectorizer. The extracted features are trained and predicted using four machine learning algorithms namely Naïve Bayes, Random Forest, Logistic Regression, SVM.

The following table shows the results arrive from our implementation model for N-gram feature extraction and prediction models.

Algorithm	Accuracy
Naïve Bayes	66.67
Random Forest	70.37
Logistic Regression	69.13
SVM	74.07

**Table: Experimental Analysis of N-gram Model**

The following table shows the results arrive from our implementation model for N-count Vectorizer feature extraction and prediction models.

Algorithm	Accuracy
Naïve Bayes	70.7
Random Forest	76.54
Logistic Regression	70.37
SVM	80.24

**Table: Experimental Analysis of N-count Model**

The following table shows the results arrive from our implementation model for TF-IDF feature extraction and prediction models.

Algorithm	Accuracy
Naïve Bayes	69.13
Random Forest	76.54
Logistic Regression	74.07
SVM	67.90

**Table: Experimental Analysis of TF-IDF model**

From the above results we can understand that Naïve Bayes model is giving good accuracy on prediction.

## References

- [1] Nitin Jindal and Bing Liu. Review spam detection. In Proceedings of the 16th international conference on World Wide Web, pages 1189–1190. ACM, 2007.
  - [2] Song Feng, Longfei Xing, Anupam Gogar, and Yejin Choi. Distributional footprints of deceptive product reviews. ICWSM, 12:98–105, 2012.
  - [3] Eileen Fitzpatrick, Joan Bachenko, and Tommaso Fornaciari. Automatic detection of verbal deception. *Synthesis Lectures on Human Language Technologies*, 8(3):1–119, 2015.
  - [4] Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T Hancock. Finding deceptive opinion spam by any stretch of the imagination. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1, pages 309–319. Association for Computational Linguistics, 2011.
  - [5] Yla R Tausczik and James W Pennebaker. The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24– 54, 2010.
  - [6] Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie S Glance. What Yelp fake review filter might be doing? In ICWSM, pages 409–418, 2013.
  - [7] Yuejun Li, Xiao Feng, and Shuwu Zhang. Detecting fake reviews utilizing semantic and emotion model. In *Information Science and Control Engineering (ICISCE)*, 2016 3rd International Conference on, pages 317–320. IEEE, 2016.
  - [8] Rupesh Kumar Dewang and AK Singh. Identification of fake reviews using new set of lexical and syntactic features. In Proceedings of the Sixth International Conference on Computer and Communication Technology 2015, pages 115–119. ACM, 2015.
  - [9] Snehasish Banerjee, Alton YK Chua, and Jung-Jae Kim. Using supervised learning to classify authentic and fake online reviews. In Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication, page 88. ACM, 2015.
  - [10] Nitin Jindal and Bing Liu. Opinion spam and analysis. In Proceedings of the 2008 International Conference on Web Search and Data Mining, pages 219–230. ACM, 2008.
  - [11] Michael Luca and Georgios Zervas. Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science*, 62(12):3412–3427, 2016.
  - [12] G. Wang, S. Xie, B. Liu, and P. S. Yu. Review graph based online store review spammer detection. In 2011 IEEE 11th International Conference on Data Mining, pages 1242–1247, Dec 2011.
  - [13] Leman Akoglu, Rishi Chandy, and Christos Faloutsos. Opinion fraud detection in online reviews by network effects. ICWSM, 13:2–11, 2013.
  - [14] Atefeh Heydari, Mohammadali Tavakoli, and Naomie Salim. Detection of fake opinions using time series. *Expert Systems with Applications*, 58:83–92, 2016.
- [1] Rehab Ashari, Charles Anderson, EEG Subspace Analysis and Classification Using Principal Angles for Brain-Computer Interfaces, Department of Computer Science Colorado State University Fort Collins, Colorado 80523
  - [2] Sarah N. Abdulkader \*, Ayman Atia, Mostafa-Sami M. Mostafa, 2015, Brain computer interfacing: Applications and challenges, HCI-LAB, Department of Computer Science, Faculty of Computers and Information, Helwan University, Cairo, Egypt
  - [3] Min-Ho Lee, Siamac Fazli, Jan Mehnert and Seong-Whan Lee, Hybrid Brain-Computer Interface based on EEG and NIRS Modalities, 1Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea 2Department of Computer Science, Berlin Institute of Technology, Berlin, Germany
  - [4] Mikhail A Lebedev, Towards a versatile brain-machine interface: Neural decoding of multiple behavioral variables and delivering sensory feedback, Department of Neurobiology Duke University Durham, USA
  - [5] Ulrich Hoffmann, Jean-Marc Vesin, Touradj Ebrahimi, Recent Advances in Brain-Computer Interfaces, Ulrich Hoffmann, Jean-Marc Vesin, Touradj Ebrahimi Signal Processing Institute Ecole Polytechnique Fédérale de Lausanne (EPFL), Switzerland
  - [6] Kup-Sze Choi, Shuang Liang, Enhancing the Performance of Brain-Computer Interface with Haptics, Centre for Smart Health, School of Nursing The Hong Kong Polytechnic University Hong Kong, China
  - [7] Chang-Hee Han, Chang-Hwan Im, EEG-based Brain-Computer Interface for Real-Time Communication of Patients in Completely Locked-in State, Dept. of Biomedical engineering Hanyang University Seoul, Republic of Korea

- [8] Tomislav Milekovic, Brain-computer interfaces based on intracortical recordings of neural activity for restoration of movement and communication of people with paralysis, Department of Fundamental Neuroscience, Faculty of Medicine, University of Geneva, Geneva, Switzerland 2 Center for Neuroprosthetics and Brain Mind Institute, School of Life Sciences, Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland
- [9] N.A. Md Norani, W. Mansor, L.Y. Khuan, A Review of Signal Processing in Brain Computer Interface System, Faculty of Electrical Engineering Universiti Teknologi Mara, 40450 Shah Alam, Selangor, Malaysia
- [10] Michael Pereira , Aleksander Sobolewski and Jose del R. Millan, Modulation of the inter-hemispheric asymmetry of motor-related brain activity using brain-computer interfaces
- [11] Junichi Ushiba, Ph.D., Asuka Morishita, M.S. and Tsuyoshi Maeda, M.S., A Task-Oriented Brain-Computer Interface Rehabilitation System for Patients with Stroke Hemiplegia, Department of Biosciences and Informatics, Faculty of Science and Technology, Keio University 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, Kanagawa, Japan
- [12] Cuntai Guan, Brain-Computer Interface for Stroke Rehabilitation with Clinical Studies, Brain-Computer Interface Laboratory Institute for Infocomm Research, A\*STAR, Singapore
- [13] Keun-Tae Kim , Tom Carlson and Seong-Whan Lee, Design of a Robotic Wheelchair with a Motor Imagery based Brain-Computer Interface, Department of Brain and Cognitive Engineering, Korea University, Seoul, Korea 2Defitech Chair in Non-Invasive Brain-Machine Interface, EPFL, Lausanne, Switzerland

