BEHAVIOR IDENTIFICATION ANALYSIS AND PREDICTION BY SOCIAL MEDIA FORENSICS

Noorulain Ashraf and Dr. Danish Mahmood

ABSTRACT

Human needs can be categorized into five levels: physiological, safety, love, esteem, and self-actualization. These needs shape human behavior, and individual behavior is influenced by events, actions, and the surrounding environment. A positive and healthy environment tends to foster healthy behavior, while an unhealthy environment can have a negative impact on behavior. With the advancements in machine learning and artificial intelligence, machines have become capable of mimicking human intelligence. In today's digital age, people frequently share their daily life events on social media through comments, pictures, videos, etc. Social media provides valuable insights into behavioral patterns and incidents over time. Studying behaviors on social media involves categorizing them as abusive, aggressive, frustrated, or offensive. Detecting a person's behavior through their social media account is an intriguing idea. However, understanding the context of the text poses a significant challenge. A person may write a post in one context, while a machine may interpret it differently. Social media forensics, which encompasses natural language processing and machine learning techniques, can help address these challenges. This proposed research critically analyzes the existing work on behavior traits exhibited on social media. Based on this analysis, a model is proposed for behavior prediction. This study contributes to the field by introducing an optimized framework for detecting criminal behavior. The proposed model will greatly benefit cybercrime and cybersecurity agencies in identifying profiles displaying specific behaviors to prevent future crimes. Psychiatrists and psychologists can also use this study to identify individuals with certain behaviors at an early stage. Furthermore, upon implementing the proposed model, it was found to outperform existing models. Five models (M1 to M5) were proposed and evaluated, with M2 demonstrating outstanding results with an f1 score of 88%.

Keywords: Behavior, Social-Media, Twitter, Forensics, Machine Learning, Natural Language Processing, Aggressive Behavior.

CARDIOVASCULAR DISEASE PREDICTION BY COMPARATIVE ANALYSIS OF ADVANCED LEARNING TECHNIQUES

Haseeb Arif and Dr.Hina Saeeda

ABSTRACT

The world is currently facing various health challenges, with heart diseases or cardiovascular diseases being a leading cause of death. Alongside cancer, these diseases account for 70% of global deaths, totaling 41 million individuals. It has been observed that these diseases are more prevalent in middle-income and low-income countries compared to high-income countries. Efficient classification of heart disease patients is crucial for saving lives. Therefore, it is essential to explore advanced learning techniques that can accurately classify these diseases with minimal attributes and testing iterations. This research aims to apply data preprocessing and regularization techniques to machine learning (ML) and deep learning (DL) models to generate realistic and unbiased results in heart disease classification. ML models such as Logistic Regression (LR) and LR with L1 and L2 regularization, as well as DL models like Neural Network (NN) and NN with dropout, have been trained and evaluated. Among these models, NN with dropout achieved the highest performance, with an accuracy of 90%, recall of 88%, precision of 87%, and f-measure of 87%.

Keywords: Machine Learning, Deep Learning, Heart Disease Classification, Cardiovascular Disease.

FAKE NEWS IDENTIFICATION USING STATISTICAL TECHNIQUES

Attar Ahmed Ail and Dr. Shahzad Latif

ABSTRACT

With the widespread use of social media, the problem of fake news has become a significant concern. Fake news dissemination can lead to various social and economic issues. Therefore, there is a critical need to develop effective methods to detect and mitigate the spread of fake news. Detecting fake news accurately is a challenging task, but it is essential for preventing its harmful effects. In this research, a model called SWELDLFake (Statistical Word Embedding over Linguistic Features using Deep Learning) is proposed to address the issue of fake news detection. The model utilizes statistical techniques, specifically principal component analysis (PCA), to extract important features from textual representations of fake news. It also incorporates word embedding to capture linguistic features. For the classification of news as true or false, a deep learning model known as Bidirectional Long Short-Term Memory (Bi-LSTM) is employed. To validate the effectiveness of the proposed model, a benchmark dataset called WELFake, comprising around 72,000 news articles collected from different benchmark datasets, is used. The results demonstrate that the SWELDLFake model achieves a classification accuracy of 98.52%, surpassing the current state-of-the-art deep learning and machine learning models in fake news detection.

Keywords: Deep Learning, Bidirectional Long Short-Term Memory, Linguistic Features, PCA.

SARCASM DETECTION IN SENTIMENT ANALYSIS USING HYBRID NEURAL NETWORK

Ahmad Hussain and Dr. Noshina Tariq

ABSTRACT

Sarcasm is often used as a means to express negative opinions using polite language, both in person and on social media. Detecting sarcasm in text is a binary classification problem that determines whether a statement is sarcastic or not. This study introduces the concept of bidirectional long short-term memory (BiLSTM) in deep learning as a method to detect sarcasm. The model employs two LSTM layers, utilizing word embeddings based on Global Vectors for Representing Words (GLoVe) in both forward and backward directions. Extensive experiments were conducted on benchmark datasets, specifically the Headlines dataset, which consists of sarcastic sentences from The Onion and non-sarcastic sentences from HuffPost. Machine learning classifiers such as support vector machines, random forest, naive Bayes, decision trees, logistic regression, and k-nearest neighbors were also explored for sarcasm detection. The highest accuracy achieved among the machine learning classifiers was 79.64% using naive Bayes. Other classifiers achieved accuracies ranging from 78.63% (logistic regression) to 57.21% (k-nearest neighbors). However, the proposed deep learning model, BiLSTM, outperformed the machine learning classifiers as well as the state-of-the-art model. The BiLSTM model achieved an accuracy of 85.08%, precision of 83.05%, recall of 87.51%, and f1-score of 85.22%. The proposed method not only demonstrated state-of-the-art performance across various datasets but also offered improved interpretability.

Keywords: Machine Learning, Sarcasm Detection, Deep Learning, Multi-Dimensional Representations, Irony Detection, Sentiment Analysis, Word Embedding.

HEART DISEASE PREDICTION USING MACHINE LEARNING TECHNIQUES

Aizhar Ali and Dr.Shahzad Latif

ABSTRACT

Cardiovascular diseases, particularly coronary heart disease, are a leading cause of death, necessitating the development of effective prediction methods to identify individuals at risk and provide timely treatment. Machine learning techniques have been utilized for heart disease prediction, but the accuracy of these methods can be limited, often due to low-quality datasets. Many researchers have explored this problem, yet the accuracy of heart disease prediction still requires improvement. An important aspect of achieving better results is the selection of significant features from the dataset. This proposed work aims to address both the quality of the analytical dataset and the accuracy of heart disease prediction by combining machine learning techniques. The NSGA technique is employed to identify significant features that play a crucial role in predicting heart disease. The proposed logistic regression technique demonstrates impressive performance on a hybrid dataset, merging the Cleveland dataset with the Z-Alizadeh Sani dataset. It outperforms existing methods for heart disease prediction. Overall, this methodology improves upon the limitations of previous studies in the prediction of heart disease. By focusing on dataset quality and utilizing advanced machine learning techniques, the proposed approach shows promise in enhancing accuracy and aiding in the timely detection and treatment of heart disease.

Keywords: Heart Disease, Heart Disease Prediction, Machine Learning, Machine Learning Techniques.

SECURING   HVAC SMART BUILDINGS FROM CYBER ATTACKS

Asad Elahi and Dr. Danish Mahmood

ABSTRACT

The HVAC system is a crucial component of smart buildings, allowing for automation and improving user comfort. However, the increased connectivity and communication in smart buildings pose cybersecurity challenges. This study focuses on data-driven attacks targeting smart buildings, particularly through the HVAC system. Two major challenges in this domain are the class imbalance problem in datasets and the vulnerability of centralized environments. Addressing the class imbalance problem is essential to avoid biased results in machine learning models. The Synthetic Minority Oversampling Technique (SMOTE) and Standard Scalar Technique are used in conjunction with Apache Spark to handle large amounts of data efficiently. Additionally, three datasets are merged horizontally, and precise feature engineering is performed to optimize the training of the learning model. Performance metrics such as Precision, Recall, and F1 score are utilized to evaluate the proposed model. Compared to existing state-of-the-art approaches, the proposed model achieves optimal results. Using the standard scalar technique, it achieves 99% precision, 95% recall, and 97% F1 score, surpassing the state-of-the-art model's performance of 90% precision, 60% recall, and 50% F1 score.

Keywords: HVAC System, Variable Air Volume (VAV), Cyber Attacks, Distributed Processing, Central Limit Attack (CLA), Centralized Systems Attack (CSA), And Single Control Attack (SCA).

SELF-SUPERVISED LEARNING THROUGH INFANT EYES

Abubakar Masood and Dr. Sohail Asghar

ABSTRACT

2022

DISTRIBUTED DATA PROCESSING AND LEANING APPROACHES FOR CREDIT CARD FRAUD DETECTION SYSTEM (CCFDS)

Taj Wali and Dr. Danish Mehmood

ABSTRACT

Credit card is a popular and widely used mode of e-payments at online stores and physical stores these days. With the invention of credit cards, payments are now more accessible and comfortable than ever. However, there are also various threats associated with these payments, and one of the most common ones is the fraud of online credit card transactions. In the modern era of big data and technology, credit cards are crucial for all stakeholders. Credit card scams are a global issue as both developed and developing countries are at serious threat due to cybercrimes. Credit card frauds have a global impact on its usage, increasing alarmingly. To date, millions of US dollars are lost by companies and cardholders with these frauds. Financial institutions lose billions of money every year to the frauds associated with credit card transactions. Fraud detection systems are used to discover fraudulent activities while the transaction is in process to handle this problem. Such systems are supported by advanced machine learning (ML) techniques and other supportive algorithms to detect and prevent fraud in the future timely. This work analyses the various ML techniques to detect fraud efficiently and explores additional state-of-the-art methods to assist with their performance. Also, this work focuses on the class imbalance issue in the credit card transactions domain. A dataset with imbalanced data creates biasness, leading to biased results at the end. The results are not accepted professionally if an ML model is trained for classification without balancing both classes. ML methods are used to handle the class imbalance of any dataset. One of these methods is the Synthetic Minority Oversampling Technique (SMOTE). This study's primary focus is on using Minority Class Clustering SMOTE (MCC-SMOTE). Simple SMOTE produces noisy data that badly affect the results. MCC-SMOTE is a clustering technique. First, it looks at the clusters of under sampled data (minority class) and then uses them to generate synthetic data. In this way, actual instances are created from similar data points. MCC-SMOTE is used with Apache Spark to handle a large amount of data in less time. Six different datasets are utilized and are processed in a distributed environment. The distributed environment uses tools like Spark for fast processing, which can handle and process a large amount of data in less time. Spark uses in-memory processing, and it works with Hadoop clusters to make the data processed 100 times faster than traditional data processing & ML methods.

Keywords: Balancing Data in Machine Learning, Balanced Bagging Ensemble, Credit Card Fraud Detection, Imbalanced Datasets in Machine Learning, MCC SMOTE Using Spark.

EARLY DETECTION OF PULMONARY NODULES ON CT IMAGE USING DEEP LEARNING

Syed Fahad Ahmed and Dr. Muhammad Imran

ABSTRACT

CT screening is a very effective method in diagnosing lung cancer. It helps in identifying pulmonary nodules at its early occurrence so to increase a patients survival chances and reduce death rate. It is a difficult and time consuming task to identify pulmonary nodules manually, especially at a very early stage. There are not many trained Radiologists and it often requires a lot

of tiring work to go through each CT slice and identify nodules, which is indeed an exhausting task. Availability of a very large image data in the form of slices makes their accurate evaluation a very challenging and time consuming task. There are various techniques which have been developed to automatically detect and classify pulmonary nodules in medical images such as those based on deep learning and convolutional neural network (CNN). However, automatic detection and classification of pulmonary nodules using CNN in a given medical image datasets is also a challenging task due to variability in lung nodule appearance and non-lung objects around them. In this research study we are going to incorporate contrast stretching image processing technique before we extract candidate nodules to be fed in to our deep learning model. Our focus is to perform and improve pulmonary nodule identification such as detection of a nodule by segmentation of morphological features using image processing and UNet model and then perform false positive reduction by classifying nodules and non-nodules using the annotated LUNA16 dataset.

Keywords: Artificial Intelligence, Machine Learning, Deep Learning, Convolutional Neural Neworks, CNN, UNet, U-Net, Segmentation, Lung Nodules, Lung Cancer, Pulmonary Nodules.

MALWARE CLASSIFICATION USING A.I TECHNIQUES

Sharjeel Riaz and Dr. Shahzad Latif

ABSTRACT

The IoT devices usage is increasing exponentially with the spread of internet. With the increasing capacity of data on IoT devices, these devices are becoming venerable for malware attacks. Malware detection is an important issue in IoT devices now a day. An effective, reliable, and time-efficient mechanism is required for the identification of sophisticated malwares. Malware detection in IoT devices is a NP hard problem. Evolutionary computing with the combination of Machine Learning techniques is good to deal with NP hard problems. To deal this issue, machine learning GA-KNN (Genetic Algo-K - Nearest Neighbors), PSO-KNN (Particle Swarm Optimization-K - Nearest Neighbors) and deep learning based algorithm hybrid CNN-LSTM are proposed. The proposed system works to disassemble the malware and extract out the byte code of the malware. A bag of word techniques is applied to develop datasets from all the disassembled files. These algorithms are simulated and compared with other state of the art machine learning and deep learning algorithms studied in literature. Simulation results showed the validity of the

proposed methodology. Moreover, results are compared with the state of art techniques studied in literature.

Keywords: Malware Classifications, AI Techniques, Internet of Things, PSO-KNN, CNN-LSTM, GA-KNN.

STANCE DETECTION IN SOCIAL MEDIA, WEBSITES TO IDENTIFY FAKE NEWS AND PROPAGANDA IN URDU

Shaan Muhammad Khan and Dr. Muhammad Imran

ABSTRACT

Widespread reach of inauthentic information has made all kinds of news suspect and trust in news has gone down drastically. The main engine through which fake news spreads on massive scales is social media. Globally, a number of deep learning approaches have been developed to detect fake news that do very well on content in English language. However, no deep learning systems are currently on the market to detect fake news/stance detection found in the Urdu language content. This presents a daunting challenge for Pakistani government authorities to counter fake news as they create problems of law and order. This research attempts to find a solution to this problem by suggesting a framework based on utilization of advanced BERT, ELMO and other deep learning models (CNN, LSTM, Bi-LSTM) to test the performance accuracy on Urdu datasets (Liar-ProSOUL and BentthTruth-Benchmark). The results of my experiment find that ELMO performs well on large datasets. Deploying our findings to Urdu content will greatly forward the technology available to detect and eliminate fake news in Urdu language.

Keywords: Deep Neural Networks, Urdu Fake News Detection, Natural Language Processing, ELMO, Deep Machine Learning.

AUTISM SPECTRUM DISORDER PREDICTION USING MACHINE LEARNING TECHNIQUES

Fatima Aslam and Dr. Muhammad Bilal Qureshi

ABSTRACT

Autism Spectrum Disorder is a severe, life-long neuro-developmental disease characterized by disabilities that are chronic or limited in the development of socio-communication skills, thinking abilities, activities and behavior. In children aged two to three years, the symptoms of autism are more evident and easier to recognize. It is important to reduce the diagnostic time and minimize the number of items used during the diagnostic process, particularly, now after the rapid growth of the smartphone industry. In the year 2020, the Centers for Disease Control and Prevention announced that according to 2016 figures, around 1 in 54 adolescents in the United States were categorized with behavioral disorders such as autism spectrum disorder (ASD). There was a classification of 1:4 in boys with autism and 1:144 in girls with autism. The prevalence rate of autism in children shows the importance that why it is necessary to predict and diagnose autism on time. It is a very critical problem for health care providers and administrators during these years of sickness, so they can develop a strategy for reimbursement and catastrophe mitigation. Many mathematical models are developed on the autism analysis, but due to the complexity of the brain, these models are unable to accurately predict autism. The symptoms of autism spectrum disorder are very similar to schizophrenia disease due to which autism prediction is very challenging and difficult. Thus, with the use of functional Magnetic Resonance Imaging (fMRI) images dataset, accurate and precise autism prediction is the main purpose of this study. The processing and extracting useful data from fMRI images can help in predicting autism spectrum disorder.

Keywords: ABIDE, Autism Prediction, ASD, Autism Spectrum Disorder, fMRI.

EARTHQUAKE PREDICTION MODEL BASED ON SEISMIC INDICATORS USING MACHINE LEARNING TECHNIQUES

Seemal Tariq and Dr. Muhammad Bilal Qureshi

ABSTRACT

This research abstract discusses the importance of earthquake prediction and the challenges associated with accurately forecasting earthquakes. Earthquakes are devastating natural disasters that cause casualties, extensive damage, and economic destruction. The abstract highlights the need to study and understand the phenomena behind earthquake prediction to mitigate these impacts. The difficulty lies in the non-linear nature of earthquake data and the inability to interpret precursor signals emitted by the Earth before major earthquakes. The abstract mentions that earthquake prediction models have the potential to reduce property damage, save lives, and enhance safety in critical facilities. Researchers have been pursuing earthquake forecasting for many years, but accurately predicting the magnitude, location, and timing of earthquakes has been challenging due to data issues. Machine learning methods, particularly neural networks, have been applied to earthquake prediction, considering the non-linear and stochastic nature of earthquake occurrence. The abstract further discusses various machine learning techniques used in earthquake prediction, including neural networks, SVM, Naive Bayes, and Random Forest. It emphasizes the superiority of artificial neural networks in generating quality outputs. The study described in the abstract combines sixty seismic indicators and applies feature engineering techniques to extract relevant features. Regression is then applied to estimate earthquake magnitudes, which are used as intermediate attributes for multiclass earthquake prediction. The resulting output classifies upcoming earthquakes into distinct magnitude classes (low, medium, high, and very high) and calculates mean square error and absolute error to evaluate the predictions. The seismic features used in the study are computed using well-known seismological concepts such as the Gutenberg-Richter law, foreshock frequency, seismic rate changes, total recurrence time, and seismic energy release. Feature selection criteria, including Information Gain and Normalization, are applied to extract relevant features. A regression-based classification model is used to predict the exact magnitudes of future earthquakes in the Chile, Hindukush, and Southern California regions. The results are evaluated using parameters such as POD, FAR, Accuracy, MCC, MSE, and MAE.
Keywords: Earthquake Prediction, Classification, Earthquake occurrence, Forecasting, Machine Learning, Magnitude, Regression, Seismic Indicators.

EXTRACTIVE TEXT SUMMARIZATION OF NEWS ARTICLES USING MACHINE LEARNING TECHNIQUES

Muhammad Mohsin and Dr. Shahzad Latif

ABSTRACT

Automatic text summarization (ATS) is gaining a lot more importance because a large volume of data is being generated at an exponential rate. Due to easy internet availability all around the world, a lot of data is being generated from places such as social networking websites, news websites, blog websites, etc., and a large portion of it is in the textual format. Users like news analysts want a summary or main crux of a news article in less time without reading the entire news article. Manual summarization is time-consuming and it is difficult to read and summarize a large amount of content. Automatic text summarization is the solution to deal with this problem. This study proposed two automatic text summarization models which are Genetic Algorithm with Hierarchical Clustering (GA-HC) and Particle Swarm Optimization with Hierarchical Clustering (PSO-HC). The proposed models use a word embedding model with Hierarchal Clustering Algorithm to group sentences conveying almost the same meaning. GA sentence ranking model is used in GA-HC for optimized ranking of the sentences. Whereas the PSO sentence ranking model is used in PSO-HC for optimized ranking of sentences. The proposed summarization models GA-HC and PSO-HC are compared with four other machine learning techniques from the literature. The results show that the GA-HC and PSO-HC techniques outperformed in comparison with other text summarization models. DUC 2007 dataset is used in this work, and Rouge 1, 2, 3, and L are used for evaluating the model.

Keywords: Extractive Text Summarization, News Articles, Machine Learning, Machine Learning Techniques.

ASPECT MINING USING DEEP LEARNING TECHNIQUES

Muhammad Owais Masood and Dr. Fadia Shah

ABSTRACT

Aspect Mining is a critical step in opinion mining and summarization task within NLP. It involves mining explicit and implicit aspects within an opinionated text. Deep Learning has gained popularity for various tasks within NLP, which include but are not limited to sentiment analysis, named entity recognition and part of speech tagging. Deep learning techniques applied to explicit aspect mining have suffered from complex network architectures causing increased processing time. Furthermore, a complex data preprocessing step, BIO encoding, is used to encode input sequence for consumption in the deep learning models. In addition to this, deep neural network hyper-parameter techniques used have been limited to dropouts and regularizations. In order to generalize well, limited research has been conducted to leverage domain specific and general embedding to generalize well across different domains and language. This research proposes a simple Convolution Neural Network based model that leverages domain specific and general embedding to generalize well across multiple domains. The model uses one hot encoding as a pre-processing step in the network architecture to reduce the simplicity of the pre-processing. Furthermore, the research provides a comprehensive framework for hyper parameter tuning including proposing a novel adaptive learning rate function. The promising results from this research are evident by an increase in the accuracy over other benchmark models.

Keywords: Natural Language Processing, Opinion Mining, Deep Learning, Aspect Mining.

2021
DEEP LEARNING ENABLED SMART DETECTION OF SOCIAL DISTANCING FOR HIGHER EDUCATIONAL INSTITUTES

Zeeshan Amin Farooqi and Dr. Muhammad Usman

ABSTRACT

The spread of the novel corona virus (COVID-19) has caused society to suffer to a great extent and to the system, to the governments and health authorities by affecting the global economies. The absence of any active therapeutic agents and the lack of immunity against COVID19 and its ability to spread in a fast manner increase the vulnerability of the population towards it. Availability of vaccines in some countries is still an issue, even after the vaccines we have to maintain the social distancing to curb the spread of virus because permanent lockdowns and curfews are not long run solutions plus. They are heavily affecting the economy as all walks of life come to a halt. As schools and educational institutes are also going to open. There is a dire need to maintain social distance. With the help of technological advancements, it is easy to monitor the violations efficiently without the human element. Hence the purpose of this study is to propose a model to detect social distancing violations in higher educational institutes. We are taking those areas where people are walking and standing. This system will detect the violations in pre-defined areas. The model which we are proposing here will do object detection on live video streams, and it will help the authorities in smooth flow of educational studies. The framework will be successfully implemented in educational institutes to detect the social distancing violations and control the spread of virus. Furthermore, we aim to improve the prediction results and efficiency of algorithms to enhance the object detection. Social distance analysis will be viewed through an analytical system which informs us about the total violations, according to our requirement; a dashboard will be integrated with the application to help us keep track of daily violations done in a day, hour or a week. It also identifies those persons who are in constant violation of social distancing.

Keywords: COVID-19, Video Surveillance, Social Distancing, Object Detection, Object Tracking.

DETECTION OF FAKE NEWS AND RUMORS ON SOCIAL MEDIA PLATFORMS

Abdul Wahab and Dr. Danish Mahmood

ABSTRACT

Alongside the development of the internet, the development and widespread adoption of the concept of social media has changed the forming and publication of news. News with social media has become quicker, cheaper and easier to view. This move is often followed by other disadvantages. Content in particular, like fake news effect users a lot and are harmful as well. Fake news can easily affect the thinking of viewer which can cause negative effects. While the fake news issue was recently introduced, it became an important subject of research because of the high content of social media. It is easy for users to write fake comments and news on social media. The main challenge is to find out how the real and the fake news differ. As in long term this fake news can affect trust of people as well over social media and internet. To keep this trust alive there should be a technique which can be fast and effective in detection of Fake news. There are many techniques in literature but most of those techniques are on downloaded data sets. In this research, a new approach is defined which can help to detect fake news on real time basis. Real time means when news is posted on twitter we can find at the same instance that how much chances are there that the news is fake or real. This is done by incorporating different factors like who posted the news? Either it is from a reputed account or any new comer. Also incorporating number of retweets for new post? For this purpose, different machine learning algorithms are used which include Logistic Regression, Decision tree algorithm, Support vector machine and Random Forest Classifier. Out of these 4 algorithms Random Forest Classifier gave best results in terms of accuracy and F1 score. This algorithm gave 0.80 of accuracy and 0.89 F1 score. Other algorithm that performed better is decision tree algorithm which gave 0.78 of accuracy and 0.80 F1 Score.

Keywords: Social Media, Fake News, Machine Learning, Kafka, Real Time, Spark, Decision Tree, Random Forest.

PRE-PRODUCTION BOX OFFICE SUCCESS QUOTIENT FORECASTING USING MACHINE LEARNING

Muhammad Idrees Hussain and Dr. Fadia Shah

ABSTRACT

The film industry generates billions of dollars annually worldwide, with Hollywood being widely recognized as the leading filmmaking hub. Despite its prominence, Hollywood has been experiencing significant revenue losses due to the failure of many films to capture audience attention in the initial days of release. Recent studies estimate that Hollywood could suffer losses ranging from 1 to 10 billion US dollars by 2020, causing immense pressure on film production stakeholders. They are constantly seeking a formula for creating successful films, but no fixed ingredients guarantee success. Researchers and filmmakers are eager to develop systems that can predict a film's fate before production, and this research proposes a decision support system for making informed film investment decisions. The study incorporates additional features such as release date, awards earned by directors and actors, Oscars, sequels, movie topics, distributor, and votes by demographics, alongside existing features. Machine learning classifiers, including SVM and random forest, will be implemented to determine the system's behavior.

Keywords: Box Office Success, Quotient Forecasting, Machine Learning.

PREDICTION OF DIABETES MELLITUS WITH EVOLUTIONARY COMPUTING TECHNIQUES

Zarnab Yousaf and Dr. Muhammad Imran

ABSTRACT

Diabetes mellitus is a chronic disease that is currently affecting many people. It occurs in a person due to the abnormal production of insulin levels in the blood. The high sugar level patients, if not treated properly, can have diabetic retinopathy, kidney failure, heart disease, and strokes. The symptoms of diabetes mellitus can be so mild that it can go unnoticed. However, it can cause severe medical issues—approximately eight million people have diabetes mellitus and are unaware of their condition. Therefore, medical data of the patient is essential to encourage the development of forecasting diabetes and subsequently, the improvement of community medicine to diagnose and treat illnesses effectively. During the period of illness, it is a crucial issue for the health care providers and managers to implement compensation and disaster-reduction plan. Small data samples are a challenge to deal with because it provides little evidence to analyze the impact of chronic diseases comprehensively. Various statistical models are developed to analyze diabetes, but uncertainty and incomplete concepts are highly related when researchers study insufficient datasets. Therefore, the main objective of this research is to predict diabetes mellitus with optimal feature selection and adaptive network fuzzy interface system model to give better accuracy. The study employed an innovative algorithm, particle swarm optimization, and mine blast algorithm continuous optimization for diabetes mellitus prediction. Feature selection will help to identify the valuable features in a dataset without losing the meaning of full dataset. The dataset used in this study contains a missing value, which needs to be controlled, and it is imbalanced because the number of subjects affected by this is significantly lower than the number of subjects unaffected. However, these findings have a significant contribution to lowering the disease burden for the subject in reducing the cost of healthcare. It also contributes to the early prediction of the disease to reduce the mortality rate. The methods and implementations for the prediction of diabetes are comprised of machine learning, artificial neural networks, fuzzy structures, genetic algorithms.

Keywords: Prediction, Model-Based Testing, Fuzzy, Diabetes Miletus, PSO, Anfis.

2021

PREDICTIVE ANALYTICS OF UNIVERSITY DROPOUT USING DATA MINING TECHNIQUES

Bilal Mehmood and Dr. Sohail Asghar

ABSTRACT

The last two decades have been most innovative in education field because of e-learning. Using that advancement and innovated way of teaching different universities started course through online service. These courses are include certificates, graduations post-graduation. On the other hand students got the opportunity of distance learning. According to the statistics of Class Central there has been in enrollment of over 101 million students around 900 universities with an estimated 11400 courses. The Open University is the largest universities in Europe while the largest in the United Kingdom. With over 2 million alumnus & 174,000 enrolled students this university offers most of undergraduate courses. Dropout is a key challenge for educational institutions in general and distance learning in specific. With a high rate of 78% dropout rate Open University is facing a challenge to overcome. Since the online courses generate a tremendous amount of data specially data from Learning Environments are generated on a daily basis. Along with sequential features there are static features which are related to student and course details. In our study we use static features along with aggregated sequential features from VLE and assessment scores. Genetic algorithm is applied in combination with Extreme Gradient Boosting to perform classification. The classification performs existing models with an accuracy of 97.46%, precision of 96.76%, recall of 98.42% & F measure of 97.58% up to week 25.

Keywords: Student Dropout, OULAD, Genetic Algorithm, Recursive Feature Elimination, Extreme Gradient Boosting.

2021
OPTIMAL FEATURE SELECTION FRAMEWORK FOR PREDICTING DEPRESSION IN USERS ON SOCIAL MEDIA USING DEEP LEARNING

Fayaz Ahmad and Dr. Bilal Qureshi

ABSTRACT

Sentiment Analysis (often called "opinion mining" or "emotional IA") refers to the use, systematically identifying, extracting, quantifying and studying of natural and affective language processing, text analysis, computer linguistic and biometrics. The sentiment classification method is extensively used to examine the user sentiments. In supervised conventional learning techniques, handcrafted functions are required, which needed in-depth domain knowledge. Since posts in social media are normally very limited and short, so in an effective classification having a lack of resources. Social media platforms represent one of the richest platforms for opinion mining giving the vast amount of user-generated data produced every day. Sentiment Analysis sometime known to as mining is the act of analyzing the publics' thoughts, feelings, reactions and emotions concerning issues and things such as services, products, policies and so on. Sentiment analysis has a lot of applications and prospects especially in the area of e-commerce, e-governance, crime combat, entertainment, and even education. For this reason, it has got tremendous significance in the community of research in recent years. In this study, this dissertation intended to develop a sentiment detection system on social media platforms bearing in mind the unique challenges presented by using informal texts. Social networking data set are valuable in many areas, including sociology and psychology. But technical support is for from sufficient and specific approaches are needed urgently. In order to detect depress user in social network services, the study apply data mining to psychology. The first is the idea of using a feeling analysis method to evaluate each microblog's depression inclination using the vocabularies and manmade rules. Second the method proposed and ten characteristic of depress user from psychology research are a depression detection model. Several models have been developed so far using different set of techniques in this regards, but still there exist a niche gape due to the paced technological advancement and availability of data. Although many state of the art algorithm for feature encoding, extraction and selection are in place in the established belt thread of machine learning paradigm and framework. Therefore, in a different context, the word embedding model can be used to learn different uses of words. To identify the polarity of feelings in short texts, using deep learning techniques, we need to find the words deeper semantics. This study explore the effects of hybrid feature space by making linear combination of N-gram, TF-IDF, and Pad sequence information and long short term memory (LSTM) , on social media for depression deprived analysis and sentiment classification. The optimum feature will be selected via random forest forward subset feature selection algorithm with rigorous 5-fold cross validation.

Keywords: Depression Analysis, Social Media Tweets And Comments Post, Long Short Term Memory (LSTM), Recurrent Neural Networks (RNN), Deep Learning, Word Embedding.

2021

OPTIMIZING YOUTUBE VIDEO POPULARITY USING MACHINE LEARNING

Meher Un Nisa and Dr. Danish Mehmood

ABSTRACT

The number of Gun related incidents have crossed over 250,000 across the globe and over 85% of the total 1 billion fire arms are in civilian hands, manual security monitoring has not proven effective to detect the firearms, which is why automated weapon detection is needed but current surveillance systems consistently underperform and several instances of attacks using weapons are reported every year. These systems require improvements that can be offered by the latest techniques and object detection models. The state-of-the art detectors are end-to-end Convolutional Neural Network based models that reach mean average precisions reaching 73%, but these models still produce a large number of false negatives in low quality videos such as, surveillance videos. This study aims to increase the accuracy in weapon detection and raise the speed of the detection because such critical systems need to detect quickly and not miss any weapon because a missed gun can cause a lot of damage and casualties. The proposed model uses You Only Look Once (YOLO) version 5s along with Area of Interest (AOI) preprocessing approach using Gaussian blur to reduce the false negative and false positive rates by blurring the background. The proposed model has achieved up to 93.8% F1 score and speeds of up to 0.010 seconds per frame in videos, which is better than the existing models used in literature so far.

2021

ECONOMY GROWTH PREDICTION OF PAKISTAN USING MACHINE LEARNING TECHNIQUES

Inaam Ur Rehman and Dr. Suhail Asghar

ABSTRACT

"Economy" is the backbone of every country. The economy of a country encompasses its trading, the products it creates and exports, and even the products a country needs to import. Everything from major trading with other countries to minor facilities available nationally, hinge on the state of the economy. This creates an emphasis for the countries to keep their state matters going in a good manner because the prosperity of their people depends highly on the economy. This is why there is a need to thoroughly understand and examine the economy of a nation and the factors affecting it. There are some indicators of the economic state of a nation which show its situation. These indicators can be used to understand how well an economy is going or running. Usually large scale and important decisions of trade are made based on these indicators. There are four main indicators of "Economy" GDP (Gross Domestic Product), EI (Employment Indicators), Consumer Price Index (CPI) and Purchasing Manager Index (PMI). We have chosen GDP for predicting "Economy Growth of Pakistan". We will predict GDP of Pakistan to determine its economy growth in coming years. The task of using a dataset containing "Pakistan Economy Data" and predicting it. It is also the novelty of this study or contribution to literature from our study. We have seen many ways of predicting economy and traditionally economy forecast is always done manually by mathematical inferences and formulas manually (without interference of modern technology) these sorts of techniques do requires a lot of work and need to involves too many experts to come up with a valuable outcome. We will use "Machine Learning" and its techniques to predict "Economy Growth of Pakistan". As we know "Machine Learning" is divided into several categories such as "Supervised Learning" and "Unsupervised Learning", furthermore these categories have more distribution as "Classification" and "Regression" for supervised learning and clustering, reinforcement for unsupervised leering. As we are using labeled data so we chose supervised learning and our problem has dependent variable a contentious one so we will not use classification method for prediction instead, we have used "Regression Analysis" such as Ridge, SVM, Lasso Regression" and we will be using optimization techniques such as Genetic algorithm and will create our fitness function for training the predictive models and will compare it results with RFE base train models.

Keywords: Economy Growth, GDP, Regression, Genetic Algorithm.

PREDICTION OF TREATMENT FAILURE IN TUBERCULOSIS USING MACHINE LEARNING TECHNIQUES

Aqib Javed and Dr. Suhail Asghar

ABSTRACT

Diabetes mellitus is a chronic disease that is currently affecting many people. It occurs in a person due to the abnormal production of insulin levels in the blood. The high sugar level patients, if not treated properly, can have diabetic retinopathy, kidney failure, heart disease, and strokes. The symptoms of diabetes mellitus can be so mild that it can go unnoticed. However, it can cause severe medical issues—approximately eight million people have diabetes mellitus and are unaware of their condition. Therefore, medical data of the patient is essential to encourage the development of forecasting diabetes and subsequently, the improvement of community medicine to diagnose and treat illnesses effectively. During the period of illness, it is a crucial issue for the health care providers and managers to implement compensation and disaster-reduction plan. Small data samples are a challenge to deal with because it provides little evidence to analyze the impact of chronic diseases comprehensively. Various statistical models are developed to analyze diabetes, but uncertainty and incomplete concepts are highly related when researchers study insufficient datasets. Therefore, the main objective of this research is to predict diabetes mellitus with optimal feature selection and adaptive network fuzzy interface system model to give better accuracy. The study employed an innovative algorithm, particle swarm optimization, and mine blast algorithm continuous optimization for diabetes mellitus prediction. Feature selection will help to identify the valuable features in a dataset without losing the meaning of full dataset. The dataset used in this study contains a missing value, which needs to be controlled, and it is imbalanced because the number of subjects affected by this is significantly lower than the number of subjects unaffected. However, these findings have a significant contribution to lowering the disease burden for the subject in reducing the cost of healthcare. It also contributes to the early prediction of the disease to reduce the mortality rate. The methods and implementations for the prediction of diabetes are comprised of machine learning, artificial neural networks, fuzzy structures, genetic algorithms.

Keywords: Prediction, Model-Based Testing, Fuzzy, Diabetes Miletus, PSO, Anfis.

2021
MACHINE LEARNING BASED SMART SITTING POSTURE MONITORING SYSTEM

Asjad Ali and Dr. Danish Mehmood

ABSTRACT

Lower backbone pain is one of the major musculoskeletal disorder that most of the individual's experience especially in common or work place. Prolonged sitting is also associated with obesity, cardiovascular risk, and premature mortality. The treatment of back problem is always difficult and most of the time it requires therapies for very long tenures. Considering the existing studies, a large number of sensors are utilized on seat pans and back rest in order to detect pressure points. However, this increase in number of sensors causes increase in the cost and complexity. Existing work also lacks in preprocessing techniques utilization for classification of sitting postures. Pervious work in literature lack in the preprocessing techniques nor encountered the optimal sensor selection and sensor placements for purposing sitting posture monitoring system. Hence, there is a dire need to orchestrate such a framework that tends to minimize the number of sensors (cost reduction- easily accessible to the masses) without any compromise on the accuracy of the verdict provided by the posture monitoring system. Therefore, the sitting posture monitoring using reduced number of sensor and optimal sensor placement is still an open challenge. The purposed model uses dimensionality reduction and feature selection technique involving PCA, High correlation filter and feature selection using Random Forest technique. The classification is performed using SVM classifier with the radial basic kernel function, using one vs all strategy. The comparative analysis with three other classifiers (Random Forest, Naïve Bayes and SVM linear kernel function) is also conducted to validate the result obtained by SVM-RBF classifier. The purposed model is designed for the classification of posture and minimize the risk of lower back pain by reducing number of sensors to four. The literature revolves around using large number of sensors to classify sitting postures, as a part of contribution our work focuses on reducing number of sensors and takes only four sensors into account to make is sitting posture classification chair economical and hit the socio-economic problem. As a part of extended contribution further the pilot study is performed to identify whether can we reduce more sensor. The pilot study also looks into different sensors combination and its effect on classification after omitting a particular sensor from the four pressure sensors used in the purposed system. Purposed model obtained the average classification accuracy ranging from 83 % - 89%. While precision ranging between 76% - 97% and have achieved Recall, and F1-Score between 87-94% and 84%-94 respectively, for all seven different sitting postures.

Keywords: Sitting Postures, Sitting Posture Monitoring System, Sensors, Sitting Posture Classification, Pressure Force Sensors, SVM Kernel Function, Machine Learning.

SYMPTOM BASED FRAMEWORK TO PREDICT SEVERITY OF HEART DISEASE

Asadullah Mushtaq and Dr. Fadia Shah

ABSTRACT

In whole world, heart related disease also known as Cardio vascular diseases are a major cause of human's death. Heart is the main organ of human body that supplies the blood in entire body. It is so necessary to predict HD along with its severity on time so patient's life can be saved. The main aim of this research is to predict actual condition of patient. In our research, we proposed a framework to predict severity of HD using Cleveland and Statlog HD datasets which are available on UCI ML repository, by applying feature selection technique PCC, Relief, mRMR and LASSO to select appropriate features which are also called as symptoms. Most used classifiers of ML like SVM, RF, KNN, DT, ANN, NB and proposed ensemble approach is applied. After getting maximum accuracy of 94.74% web-based application has been created to get the values of temperature, humidity, human perception and pulse rate along with other symptoms to calculate the HD along with severity from live data of patients using flask. Real time HD prediction application is deployed on cloud platform.

Keywords: Machine Learning, Heart Disease, Feature Selection, Support Vector Machine, Random Forest, Decision Tree, K-Nearest Neighbor, Artificial Neural Network, Naïve Bayes, Flask.

2021

AUTOMATED WEAPON DETECTION AND CLASSIFICATION WITH THREAT ASSESSMENT USING DEEP LEARNING

Abdul Hanan Ashraf and Dr. Muhammad Imran

ABSTRACT

The number of Gun related incidents have crossed over 250,000 across the globe and over 85% of the total 1 billion fire arms are in civilian hands, manual security monitoring has not proven effective to detect the firearms, which is why automated weapon detection is needed but current surveillance systems consistently underperform and several instances of attacks using weapons are reported every year. These systems require improvements that can be offered by the latest techniques and object detection models. The state-of-the art detectors are end-to-end Convolutional Neural Network based models that reach mean average precisions reaching 73%, but these models still produce a large number of false negatives in low quality videos such as, surveillance videos. This study aims to increase the accuracy in weapon detection and raise the speed of the detection because such critical systems need to detect quickly and not miss any weapon because a missed gun can cause a lot of damage and casualties. The proposed model uses You Only Look Once (YOLO) version 5s along with Area of Interest (AOI) preprocessing approach using Gaussian blur to reduce the false negative and false positive rates by blurring the background. The proposed model has achieved up to 93.8% F1 score and speeds of up to 0.010 seconds per frame in videos, which is better than the existing models used in literature so far.

Keywords: Automated Weapon Detection, Classification, Threat Assessment, Deep Learning.

2021
BRAIN DECODING USING FMRI IMAGES FOR MULTIPLE SUBJECTS THROUGH DEEP LEARNING

Laraib Azad and Dr. Muhammad Bilal Qureshi

ABSTRACT

Substantial information related to the person cerebral conditions can be decoded through various non-invasive evaluating techniques like fMRI. Exploration of the neuronal activity of the human brain can divulge the thoughts of a person like what is the subject perceiving, thinking or visualizing. Furthermore, MVPA based deep learning techniques can be used to decode the multifaceted patterns of the brain in response to the external stimuli. Existing techniques are capable of exploring and classifying the thoughts of the human subject acquired by the fMRI imaging data. fMRI images are the volumetric imaging scans which are high dimensional as well as require a lot of time for training when fed as an input in the deep learning network. However, the hassle for more efficient learning of highly dimensional high-level features in less training time and accurate interpretation of the brain voxels with less misclassification error is needed. In this thesis, we propose an improved CNN technique where features will be functionally aligned. The optimal features will be selected after dimensionality reduction. The high dimensional feature vector will be transformed into low dimensional space for dimensionality reduction through auto adjusted weights and applying strides. Furthermore, we solve the problem of increased training time by increasing the width of the network model and using Swish activation function, making it denser and increasing efficiency of the model in less training time. Finally, the experimental results are evaluated and compared with other classifiers which demonstrated the supremacy of the proposed model in terms of accuracy.

Keywords: fMRI, CNN, Deep Learning, Swish, Brain Decoding.

2020

PREDICTING UNDER-FIVECHILDMORTALITY FOR MATERNAL, NEWBORN AND CHILDHEALTH (MNCH) OF PAKISTAN

Shoaib ul Din Memon and Dr. Fadia Shah

ABSTRACT

In developing countries, improving child health and reducing under-five child mortality are top priorities. UNICEF adopted Sustainable Development Goal 3 (SDG3) for reduction in the under-five child mortality rate globally to 25 deaths per 1000 live births. Under-five mortality rate is 69 deaths per 1000 live births in Pakistan according to Demographic and Health Survey (2018). Some of the challenges in mortality reduction are accurate and timely estimation of the under-five child mortality, equitable coverage, and skilled delivery. Predictive analytics has the power to transform the healthcare industry, personalizing care to every individual. This study contributes some values in the improvement of childhood health by analyzing the child mortality risk factors by using machine learning techniques. Pakistan Demographic Health Survey (2017-2018) publicly available dataset is used in this study and multiple imputation method is adopted for treatment of missing values. The information gain, feature selection method ranked the information-rich features and examine their impact on child mortality prediction. Divide the data into training set and testing test. Synthetic minority over-sampling method (SMOTE) balanced the training dataset and four supervised machine learning classifiers have been used namely decision tree classifier, random forest classifier, naïve bayes classifier, and extreme gradient boosting classifier. For comparative analysis accuracy, precision, recall, and f1-score have been used. Eventually, a predictive model is build that predicts whether the child is alive or dead. Number of children five and under in household, Preceding birth interval, Births in last five years, Family members, Mother age, Age of mother at first birth, Antenatal care visits, Birth order number, Breastfeeding, Region, Total children ever born, Child size at birth, Place of delivery, Wealth index, and Sex of child were found to be critical risk factors for child mortality. XGB classifier performed efficiently and predicted under-five child mortality with Accuracy (0.882), Precision (0.977), Recall (0.896) and F1-Score (0.935). The result obtained could support child health intervention programs in decision making.

Keywords: Child Mortality, Developing Countries, Health Care, Predictive Analytics, MNCH.

2020

PREDICTING COGNITIVE ABILITY OF ELDERLY PEOPLE USING MACHINE LEARNING

Shoaib ul Din Memon and Dr. Danish Mehmood

ABSTRACT

The rapid increase in the world's population has led to the "silent" epidemic of our century, dementia. Over the last decade, increased attention has been paid to precursors to Dementia disease. Dementia and cognitive impairment are the leading causes of illness and the need for treatment for the elderly. Surprisingly, despite its responsibility, much less research has been done on dementia studies. The current research objective is to predict whether the subject is CN (Clinical Normal), having MCI (Mild cognitive impairment) or Dementia and when the progression of MCI to dementia occurs within four years by examining the socio-demographic, clinical, and genetic factors. The findings play a significant role in finding treatments for Dementia modification and early prediction. There is currently no treatment available for Dementia-modification. The author used the Alzheimer's disease Neuroimaging Initiative (ADNI) data collection containing socio-demographic, non-clinical assessment, and genetic data for this study. The data used in this study include missing value, which needs to be handled. It is imbalanced because the number of subjects affected by this is significantly lower than unaffected. So, Synthetic Minority Over-sampling Technique (SMOTE) did balance the data, which removed the minority problem. A grid search approach is used to find the best classification and best hyper parameter by evaluating Random Forest, Decision Tree, Gaussian naive Bayes, KNN, Support Vector Machine, and Logistics Regression techniques multi-classification. The higher the value of the F1 score, the better the accuracy. The study has shown that the support vector classifier performs sustainably better than the rest of the classifiers with the F1-score of 0.78. Also, the model is able to predict the progression but the model is insufficient to predict the moment of progression. The genetic factor provides a great impact and improves the overall result on the multi classification problem with the F1-score of 0.84. The study has succeeded in identifying the impact of genetic factors which contribute to the process. However, these findings can contribute to lowering the disease burden for the subject and reducing healthcare costs, and finding the Dementia modifying therapies, as there is no therapy available

Keywords: Mild Cognitive Impairment, Dementia, Multiclass Classification, Moment of Progression, Imbalanced Data, Decision Tree Classifier, Support-Vector Classifier, Stochastic Gradient Descent, Logistic Regression, Artificial Neural Network.

2020

AN AUTOML FRAMEWORK FOR PREDICTIVE ANALYTICS OF SMART CITY APPLICATIONS

Nazia Nawaz and Dr. Azhar Mahmood

ABSTRACT

Machine learning is a labor-intensive process and makes this process easy; automating machine learning concepts emerged as a hot topic. AutoML attempts to accelerate the process of evolving a trained model by automating the majority of machine learning steps. Users are expected to upload the datasets and wait for the predictions to become available. AutoML makes data scientists more productive as most of the steps of machine learning are automated. Similarly, the analytics of smart city applications are important to improve the services and quality of life.Smart Cities can become truly smart through data and analytics. AutoML is a tool that can help companies to get insight from massive and high-velocity big data of smart city applications. Some examples of AutoML Frameworks include Auto-Weka, Auto-Sklearn, Tpot, Recipe framework. These frameworks work for offline data so there is a need of a tool that can predict on streaming data. In this study, we proposed a Real-Time Predictive Analytical framework (RPA) for predictive analytics of offline and live data of smart city applications. The RPA is a scalable distributed framework that has been implemented using big data management tools Apache Kafka, Apache ZooKeeper, Apache Hadoop, Apache Spark, Apache HBase, and Flask. RPA selects important predictors using Hybrid Filter-wrapper optimization using the Genetic Algorithm to select the optimal feature set. The RPA framework is validated on cross domain datasets e.g. orthopedic disease data smart parking data and GPS Trajectories data. The experiment results indicate that the RPA makes a prediction on orthopedic disease data with 94% accuracy, smart parking data with 83% accuracy and GPS Trajectories data with 98.2 %accuracy. After selecting the important predictors the training time has reduced. Additionally, the RPA has a restful web service deployed using Flask to serve requests coming from user applications i.e. mobile phone.

Keywords: Automated Machine Learning, Real Time Predictive Analytics, Internet of things, Smart City, Big Data, Web Service.

2020

FRAUD DETECTION IN CREDIT CARD ONLINE TRANSACTIONS USING MACHINE LEARNING

Awais Khan Mughal and Dr. Danish Mahmood

ABSTRACT

Credit card fraud detection is currently the most prevalent issue in the world today. This is attributed to a spike in both online and e-commerce channels. Credit card theft usually occurs when the card has been stolen for some illegal reason, or even when the fraudster utilizes credit card details for individual benefit, the information on the card is actually being used to purchase goods. We're facing a lot of credit card challenges in the world today. A payment card fraud monitoring scheme has been put in order to track fraudulent practices. A large number of models were being proposed in earlier research focused on single and hybrid machine learning algorithm-based models, but each model has its own unique limitations due to biases of the dataset. In earlier researches, machine learning algorithms like Random Forest, Logistic Regression and K-Nearest Neighbor have been used in both single and hybrid models i.e., different combinations but each research has unique limitations like low rate of fraudulent transaction identification, non-consideration of user behavior and lack of considering truly prediction rather than prediction. In this research, we proposed a hybrid model based on the combination of logistic regression and a random forest learning algorithm with a gradient boosting technique. The research focuses on eradicating the limitation of earlier research by considering both accuracy and true positive rate calculation parameters along with consumer transaction behavior. The results obtained through the proposed model are quite good as the model has accuracy of 99% and a true positive rate of 98%. The results are achieved by balancing the highly imbalanced dataset by using SMOTE technique.

Keywords: Automated Teller Machine (ATM), Transactions, Machine Learning (ML), ANN (Artificial Neural Network), Deep Learning (DL), Random Forest, Accuracy, K-Nearest Neighbor (KNN), Gaussian Bayes, False Positive Rate (FPR), TPR (True Positive Rate).

DEMAND FORECASTING IN SUPPLY CHAIN MANAGEMENT USING DEEP LEARNING

Nameer Ul Haq Qureshi and Dr. Danish Mahmood

ABSTRACT

Demand forecasting is one of the essential aspects of supply chain management, as it is linked with the financial performance of the organization. For retail businesses, it is more challenging to keep a record of their products to meet customer needs. In the retail industry, it is essential to have more accurate forecasting to make suitable decisions; therefore, the selection of the right forecasting method is considered vital and ideal. The correct forecasting helps in producing the right decision, which gives sound financial output to the company. Previously, many researchers have been working on developing various models to forecast sales demand to improve accuracy. In the previous researches, it has been noticed that few of the necessary conditions or situations were not being catered to in sales demand forecasting such as most researchers used univariate datasets; considering only id, date, and total sales for forecasting demand. Internal and external factors such as weather, promotional activity, location of the store, and holidays also play one of the primary roles when it comes to sales demand to forecast which are common for all products. Therefore, it is not specifically a univariate problem but a multivariate problem which have been analyzed in this research. In this research, multivariate dataset including weather variables, other important features have been used in predicting sales demand in supply chain management which helped to achieve better and reliable results. A SDFW model is proposed using GRU with Grid search on the comprehensive data which incorporates weather data with sales data. Various pre-processing techniques were applied to format and prepare the data for the proposed model. The GRU with Grid Search showed better performance on data set with additional features with Mean-MAE 325.21, as compare to LSTM model. These findings may help the sales industry to predict the upcoming sales demand in a more efficient way in order to allow them to optimize their inventory.

Keywords: Sales Demand to Forecast, Heuristic Algorithm, Hyperparameter, Retail Business, Supply Chain Management, Recurrent Neural Network, Deep Learning.

MORTALITY RISK PREDICTION IN PCI (PERCUTANEOUS CORONARY INTERVENTION) USING DEEP LEARNING TECHNIQUES.

Junaid Hussain and Dr. Qamar Abbas

ABSTRACT

Percutaneous coronary intervention (PCI) is a crucial medical procedure for heart patients. The success of a PCI procedure relies on thorough preliminary evaluations conducted by medical experts. Patient condition and procedure severity must be critically assessed to ensure a successful surgery. While relying on the recommendations of medical experts may be satisfactory to some extent, there is always room for improvement, particularly in cases where the patient's life is at stake. The advancements in artificial intelligence and data analytics have made it possible for technology to make critical decisions based on existing data. In the field of medical sciences, mortality prediction plays a significant role in predicting clinical outcomes. The healthcare industry generates a massive amount of data on a daily basis from hospitals, medical institutes, and clinics. However, this data is often unstructured and needs to be properly organized and cleansed for predictive tasks. The structured data can be utilized for classification and prediction tasks, which in turn can aid in mortality prediction. In this study, we propose the use of a Deep Neural Network (DNN) for mortality prediction in clinical procedures. The experiment is conducted on a publicly available research dataset. Standard parameters such as accuracy, precision, recall, and f1-score are calculated to evaluate the performance of the proposed system. These parameters demonstrate the capability of the proposed model in predicting mortality.

Keywords: Mortality Risk, Prediction, PCI, Deep Learning Techniques.