



*Analytics
Society of India*

National Conference
on
Machine Learning and Artificial Intelligence
August 26-27, 2019

Conference Proceedings

Organized by
Department of Computing
Coimbatore Institute of Technology

In association with
Data Centre and Analytics Lab, IIM Bangalore
&
Analytics Society of India

LUNG NODULE DETECTION IN CT IMAGES USING CNN – NOVEL APPROACH

Preethi Bhaskaran

Assistant Professor, RSET

preethiforever@gmail.com

Dr. Jayanthi V. S

Professor, RSET

jayanthivs@rajagiritech.edu.in

Gilu K Abraham

Student, RSET

gilukabraham@gmail.com

Deep learning technique have permeated the entire field of medical image analysis because of its self-learning ability and deep feature extraction capability. The hierarchical architecture of convolutional neural networks paves their way to biomedical applications. Deep learning based lung nodule detection models have been proposed to assist radiologists to reduce their hamper on screening pulmonary nodules. The white color spot with maximum intensity present on lungs, different from normal lung tissues, visible on X-ray or Computed Tomography (CT) scan images were treated as lung nodules. This paper introduces a newly designed serial Convolutional Neural Network (CNN) for the early detection and diagnosis of lung nodules. The proposed neural network architecture was trained on LIDC-IDRI dataset. A comparison of the proposed neural network and the pre-trained VGG16 and Alexnet has been made using the accuracy metrics. A comparative analysis was also made regarding the number of layers, convolutional kernels used in each layers, input size of the image, processing time and performance efficiency. The proposed system gives an accuracy of 96.44% in 23 minutes using single CPU.

Keywords: *Deep learning, Convolutional neural network, Feature extraction, Computed Tomography.*

DEEP LEARNING ENHANCED ALZHEIMER'S DISEASE CLASSIFICATION

Blessy C Simon

Department of Electronics and Communication
Rajagiri School of Engineering and Technology
Kochi, India
blessysimon84@gmail.com

Dr. Jayanthi V S

Department of Electronics and Communication
Rajagiri School of Engineering and Technology
Kochi, India
jayanthivs@gmail.com

Alzheimer's Disease is the most common form of dementia which initially destroys the memory and finally progresses to death. This irreversible disease is mostly found among older people. The latest innovations on the multimodal neuroimaging data made it possible to detect the disease in life which was a major breakthrough in neuroscience. However, the larger degree of similarity between the brain images was the major challenge in the diagnosis. The Deep Learning technique has gained excellent results on image classification among the present researches. Hence it is utilized for the classification of brain images among Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), Mild Cognitive Impairment (MCI), Late Mild Cognitive Impairment (LMCI), Alzheimer's Disease (AD) which are the five classes of AD thus ensuring very precise and accurate diagnosis. The transfer learning approach has been taken up for the classification process by which three pre-trained networks, namely AlexNet, ResNet-18 and, GoogLe Net are modified and trained for 180 images. All the three networks are trained for the same set of images which were acquired from the ADNI database.

Keywords: *Alzheimer's disease, Deep learning, Classification, Convolutional Neural Network, Transfer learning, Pre-trained Networks.*

DATA ANALYTICS WITH BLOCKCHAIN

Sudendar S*, **Karthikeyan S***, **Jeeva K.G***, **Anesh P***, **Kamesh G***, **Sudha Devi D#**

* III year M.Sc. (Data Science), Department of Computing

Associate Professor, Department of Computing

Coimbatore Institute of Technology,

Coimbatore, Tamil Nadu, India

sudhadevi.cit@gmail.com

Blockchain technology is poised to change nearly every facet of our digital lives. Blockchain can be said as an immutable, and decentralized database. Data stored in the blockchain cannot be tampered, making it secure. Also, being decentralized, no central entity controls the blockchain, ensuring reliability. So, the data can be stored publicly, such that anyone could read the data. This vast availability of data could enable data scientists to perform various analytics over the large amount of data. This could result in many useful insights in many fields. But, when sensitive data such as healthcare data and reports are to be stored in the blockchain, it could raise several privacy issues. Medical reports or personal information cannot be stored in a way that anyone could access them. Thus, this paper suggests a way to store and perform analytics over sensitive data in blockchain. In this paper homomorphic encryption is used to store the sensitive data in blockchain. Homomorphic encryption allows computation on ciphertexts, which gives an encrypted result. When the encrypted result is decrypted, it gives the result that matches the result of the operations as if they had been performed on the plaintext. Thus, critical data are encrypted and stored in the blockchain, data analytics are performed over them, without knowing the actual data. Thus, the proposed system provides privacy of the data stored publicly and could also profit the data scientists with access to large amount of real time data directly from the owners of the data.

Keywords: *Blockchain technology, homomorphic encryption, healthcare data, data analytics.*

USE OF DATA VISUALIZATION FOR VARIANCE ANALYSIS IN DISCRETE MANUFACTURING

Dilip Kumar R

ravichandrandilip@gmail.com

Gowtham Balaji P

gowtham2balaji@gmail.com

Hariprashanth M

hariprasanthmadhavan@hotmail.com

The broad range of manufacturing components for given set of products in a discrete manufacturing industry has resulted in the conception of a complex system of management, in terms of financial costing calculations. The raw material items for a product, range from minor parts to huge mechanical structures. On a constant basis, the manufacturers have the need to identify, if the cost has increased or decreased for a given cost structure. The criticality lies in the correct measurement of the ratio of Price and Quantity variance and its contribution towards the overall Cost of the product. The understanding of these contribution factors will aid the company in understanding the existing problems, defining the scope for course corrections and formulating new strategies. It is important how to effectively use the Data utilization tools, to quickly understand and dig deeper, if the cost variance is caused by the change in Price or Quantity of a given item. Thereby, helping to gather insights on the underlying complex problems for financial costing team, inside a discrete manufacturing company.

Keywords: *Data Visualization, Financial Costing, Variance Analysis, Discrete Manufacturing.*

RETAIL ANALYTICS – MARKET DOMINATION

C R Deepak Kumar

3rd semester, Integrated MSc. (Decision & Computing Sciences),
Department of Computing Sciences,
Coimbatore Institute of Technology,
Coimbatore, Tamil Nadu, India.

crdkapri13@gmail.com

We all want to know ‘Why is this company able to dominate its particular sector?’, while there are other companies who sell the similar kind of goods at a cheaper price, also the quality of the product is bearable and yet the company who sells at a higher price still dominates. I think there is an answer to ‘why they succeeded?’ in every success story. We can also understand then why are they the ones still dominating the market and not us since we make the same products which are at par with the quality level standard and also cheaper. I think all of this boil down to the drive of a particular organization or a team. Simon Sinek clearly in his book states if your product connects with the audience then you will surely win the market. Why the drive of the company works, Biology will say how.

The present research will suggest companies to look at the combination of ‘Reengineering managerial approach and Management Science approach’ for becoming the leading managerial way to dominate the marketing sector. The study will give insights to various intellectuals who are planning a startup to get early domination at the market with their products. The Analysis and survey will be conducted to prove to you that this should be how business should operate to acquire more profits.

Key words: *Psychology, right customers, reengineering managerial approach, management science approach.*

INTELLIGENT WAREHOUSE MANAGEMENT SYSTEM WITH MACHINE LEARNING

**Ashish Majumdar, Debdut Saha, Sharmistha Sanyal, Vineela Allamneni
Janice Ann George, Raja Krishnamoorthy, Arup Ray**
Cognizant Technology Solutions, India

In this paper authors investigate emerging trend of applying machine learning algorithms in solving dynamic business challenges in a large & complex warehouse. While traditional warehouse management software-s in combination with barcode scanners, RFID and AGV-s had been successful in automating transactional processes, machine learning is bringing new opportunities to introduce intelligence to warehouse management system replacing manual intuition and experience based approach. In this case study, machine learning algorithm has been deployed to identify Fast, Slow & Non-moving (FSN) products in a dynamic environment to arrive at an appropriate placement strategy leading to significant improvement in efficiency & productivity. This paper explores the solution approach using Machine Learning for the warehouse management processes leading to optimal usage of resources.

Keywords: Warehouse Management, Machine Learning, Big Data, Tensorflow, Python.

CHATBOT –METADATA DRIVEN

**Mathi Venkatachalam
Gunasekaran Arumugam**

Chatbot is a conversational agent that can interact with end users in a given subject using the Natural Language Processing (NLP). Many Chatbots have been deployed on the internet for the purpose of education, customer service, guidance, entertainment and goes on. Over the last few years, chatbot have played a prominent role as human-computer interfaces. Chatbots are generally composed of three modules: the user interface, an interpreter, and a knowledge base. In this paper, we talk about meta-data driven chatbot framework which allows the chatbot team to define, design, build and deploy the chatbot in short period of time (less than half a day). The framework that allows the team to build bots without coding skills. And every message delivered to the user and every action performed by bot is completely controlled by the information defined in metadata framework.

Keywords: *Chatbot, NLP, Metadata*

PREDICTING SUCCESSFUL HIRES USING MACHINE LEARNING

Sheethala M

III year, M.Sc., Data Science,
Department of Computing
Coimbatore Institute of Technology, Coimbatore, India
msheethala@gmail.com

Keerthana K

III year, M.Sc., Data Science,
Department of Computing
Coimbatore Institute of Technology, Coimbatore, India
krishikeerthu@gmail.com

Recent trends have changed the candidate recruitment process. It's a candidate driven market out here. Compared to a few years ago, candidates now have far more power during the job search. When there are more job seekers looking for work than jobs available it's becoming common for candidates to reject a job offer. Candidates refusing job offers is rarely out of company's hands completely. According to research study and every recruiter and HR professional everyday work experience, the current job market is 90% candidate driven. That means company don't pick talent anymore, talent picks company.

Consequently, finding and hiring perfect job candidates, especially those with in-demand skills, has become an extremely hard, expensive and time-consuming. No one wants to waste time with rejected offers. In recent trend, candidates get multiple job offers from different companies and give the false promises to the recruiter leaving them to guess their joining status.

The aim of this paper is to predict which candidates are more likely to accept a job offer based on predictive modelling and machine learning (ML). The gathered dataset is about prior candidates who have and have not accepted such offers. By using this model recruiters can make more informed decisions about the likelihood of candidate success, thus bettering their overall rate of successful hires. Predictive models can be trained over time to respond to new data, delivering the results to HR needs. The predictive model is being built using classification algorithms like neural networks and random forests, which fall in the category of Supervised Machine Learning.

Keywords: *Predictive Models, Machine Learning, Supervised Learning, Job Acceptance, Recruitment.*

INFLUENCER MARKETING AND CONSUMER BUYING BEHAVIOUR: USING SUPERVISED MACHINE LEARNING ALGORITHMS

Dr. Biranchi Narayan Swar

Associate Professor

Symbiosis Institute of Business Management

Symbiosis International (Deemed University)

Bangalore, India

drbiranchi.marketing@gmail.com

Anshu Srivastava

Research Assistant

Symbiosis Institute of Business Management

Symbiosis International (Deemed University)

Bangalore, India

anshu.srivastava@sibm.edu.in

Over the past decade, Influencer marketing has evolved from traditional marketing strategies such as print ads, celebrity endorsement to digital marketing. To gain the competitive advantage, retailers are looking at consumers' insights. Hence, the purpose of this paper will focus on analysing the impact of influencer marketing on consumer buying behaviour. This paper will conduct both qualitative and quantitative approach to analyse the data. A sample of 201 respondents will be considered for the survey and will be analysed using robust machine learning technique like predictive analytics modelling. This study will contribute to the literature by validating influencer marketing's impact on consumer buying decision. This is to get a better understanding of what is needed to be taken into consideration before starting a collaboration between a retailer and an influencer. The paper will bring originality in presenting a fresh attempt to influencer marketing, and its relationship with consumer buying behaviour.

SPEECH NOISE SUPPRESSION USING DEEP LEARNING

Mr. Velmurugan Munusamy

Research Scientist

Logituit

Coimbatore, India

Velmurugan.munusamy@logituit.com

Mr. Srinivasan Sampath

Solution Architect

Logituit

Coimbatore, India

Srinivasan.sampath@logituit.com

Nowadays the use of speech signal data for communication in real time, speech-noise removal is more important for clear communication. The speech signal, that is recorded by any sources like phone is generally infected by noise originating from various sources like Dog barking, traffic sound, horn sounds and various environment noises. These noises can change and affect the quality of speech signal. So, we present a deep learning algorithm for speech-noise removal. Given input audio containing speech corrupted by background noise signal, the objective of this paper aims to produce only the pure speech content without background noise. Recent approaches have shown good results using various deep learning approaches. In this paper we proposed modified RNN for speech noise removal. Through the experiments our proposed approach can perform better than some deep learning approaches and that it is able to remove noise well on speech signal across various noise types.

Key Words: *Noise Suppression, RNN, Deep learning.*

PREDICTABILITY OF GLOBAL STOCK MARKETS - A NEURAL NET ANALYTICS

Thirupparkadal Nambi S

D.J. Academy for Managerial Excellence,
Pollachi Main Road, Othakkalmandapam,
Coimbaore- 641032, Tamilnadu.

nambist@gmail.com

Deepa Venugopal

D.J. Academy for Managerial Excellence,
Pollachi Main Road, Othakkalmandapam,
Coimbaore- 641032, Tamilnadu.

getdeepadeep@gmail.com

Subha M V

Anna University Regional Center, Maruthamalai Main Road,
Navavoor, Coimbatore-641046, Tamilnadu.

subhamv@gmail.com

As stock market is a real-time, complex, dynamic, and vibrant system, it poses a challenge to the academia to unravel its dynamics and functions. On the other hand, investors are lured by the prospects of making fortune with the ability to correctly predict the price and the trend of the stock market. The knowledge of market dynamics is of immense value to individual investors, institutional investors, financial institutions, and others whose fortunes are largely dependent on the movement of the stock market. It is also useful for regulatory bodies to formulate suitable policies for efficient functioning of the system. Stock market indices reflect up-to-date information and hence they are considered as lead indicators of the performance of the overall economy. The current research aims at investigating the predictability of the major global indices using neural network. Neural Net is an efficient information processing system that resembles the characteristics of a biological neural network. It possesses a large number of highly interconnected processing elements called nodes. Each connected link is associated with weights that contain information about the input signal. This information is used by the neural net to solve a particular problem. The daily closing price of 31 major global indices, including their intraday movements such as open, high, low, close price along with the daily volume of transaction and percentage of change, for the period of three years ranging from 1st January 2016 to 31st December 2018, comprising more than 800 trading days, are considered for this study. Results reveal that the best time horizon for prediction is found to be daily as the daily prediction errors are significantly less compared to weekly and monthly prediction errors. S&P BMV IPC of Mexico is the less predictable market with the relative root squared error of 20.4% and Dow 30 is found to be the highly predictable market with the error of just 5.08%. The correlation coefficient between prediction error and market returns are negative (-0.6) implying that it is easier to predict performing indices.

ADAPTIVE INTELLIGENCE BASED LONG RANGE FORECASTING (AiBLRF)

An Overview of Rule-Based Prediction Method

Gyan Prabhat

Statistical Analyst

Walmart Labs

Bangalore, India

Gyan.Prabhat@walmartlabs.com

Pinaki Brahma

Data Scientist

Walmart Labs

Bangalore, India

Pinaki.Brahma@walmart.com

One of the major requirements for any forecasting technique is to have historical data for the past few years to capture the trend/seasonality and other parameters. The number of data points required to build any statistical model depends on two things: the number of model coefficients (parameters) to estimate and the amount of variance in the data. Statistically, it is always necessary to have more observations than parameters.

This brings us to the question, what if we don't have such amount of data?

In this report we will present a use-case of how logical rules can be helpful for such cases. Sales in countries like China are majorly influenced by the lunar festivals and these festivals occur in varying dates in different years. When there is not an abundance of data, to get the estimates of the impact of such lunar holidays might be tricky and obviously, one can't keep waiting to first accumulate few years of data and then use the conventional forecasting techniques.

The business knowledge of the prior and post impacts of different lunar holidays could be used to achieve a reliable forecast. A set of rules have been formed after hours of discussions with business. Although not every time are these rules statistically justifiable, but these rules are always concrete, perceptible, tangible and yet sufficiently flexible. As these rules have been formed using business knowledge, it develops a trust factor for the business on the forecast model.

The business does not want to have only trust-able forecast values but also a fair understanding of how these forecast values have been calculated, so that it can be adjusted if necessary. Thus, in such cases where the how of the forecast is not visible to the business (like a black-box approach), it gets thrown off the shelf. Also, business will need the flexibility to have control on the impacts of holidays. If they have any other knowledge about those holidays, they might change the calculated impacts.

In this paper, we will give a comparative study of how the rulebased system performs better in such cases, compared to the other statistical forecasting methods.

Keywords: Lunar holidays, Rule-based method, Impact of holidays.

INCREASING ACCURACY OF SPAM TEXT CLASSIFICATION USING FEATURE ENGINEERING

Gyan Prabhat
Statistical Analyst
Walmart Labs
Bangalore, India
Gyan.Prabhat@walmartlabs.com

The amount of Textual Information being generated has increased drastically at present and with it comes various problems related to Natural Language Processing (NLP). One of the basic use cases of NLP and text-analytic is to identify spam texts. Manual identification of spam text is time consuming and tedious and thus automated approaches are required, and these automated processes should have good accuracy and recall.

Given a set of legitimate texts and spam texts, where there is high imbalance between the two classes, the goal is to identify the spam texts with higher accuracy. In this paper we will try to present a comparative study of different classification methods to identify spam texts. Then we will discuss the use of GloVe method to identify the commonly occurring word-pairs in spam texts and will use this information to form a metric. This metric will be formed using the key words identified earlier, and will consist of two unique components, absence or presence of the key words and the frequency in which those key-words are appearing in a text. This metric will be used to assign a score to each text and this will be a new feature to use in the classification method.

At the end, it will be shown that using this metric which was formed by identifying the commonly co-occurring words in the spam texts can increase the accuracy of the classification method.

Keywords: *Text classification, Co-occurring words, GloVe, Feature Engineering, Accuracy.*

EFFICIENT HARDWARE ARCHITECTURE FOR AI INFERENCE

Aditya Sivagnanam

Senior FAE

Xilinx Development Corporation

Bengaluru, India

Aditya.s@xilinx.com

Sundar Krishnaraj

Staff FAE

Xilinx Development Corporation

Bengaluru, India

sundark@xilinx.com

With more and more AI inference applications becoming mainstream there is a requirement that the underlying silicon run more efficiently both at the edge and in the cloud. Often the edge sensor has a very low power budget and that demands a power efficient architecture for the compute demands from DNNs. In this article we present a solution for power efficient architectures that are non-traditional (non CPU/GPU) silicon devices and are best suited for inference workloads. These are called DSA or Domain Specific Architecture that sculpt the silicon device to exactly match the model compute requirements. In addition to being power efficient, this also makes the inference engine a very low latency function. Applications like autonomous driving, video analytics or speech-to-text immensely benefit due to the low latency it can offer. The GPUs or CPUs are fixed silicon architecture that is designed for a generic workload in a typical compute or graphics application. It offers only fixed datatypes that any representation should fit in. Several studies [1] in the past have shown that DNN models doesn't need floating point precision at inference, an INT8 or INT2 precision is enough to achieve similar or no loss in accuracy. This requires that a DNN processor should offer varying precision to make it efficient in area, power and performance. The solution presented here is based on a SoC architecture with an in-built Hardware Accelerator that is reconfigurable (FPGA – Field Programmable Gate Array). The SoC is built using massive arrays of digital gates floating in a network of programmable interconnect alongside a processing system. The compute intensive math functions like matrix multiplication, vector multiplication used in DNN layers are accelerated using the reconfigurable accelerator logic. It also allows flexibility to parallelize these functions to achieve the desired performance, latency and power efficiency. Xilinx offers several pre-canned hardware overlays called as DPU or xDNN that provide hardware acceleration to the processing subsystem within the SoC or to the host CPU. This article presents this innovative architecture and provides comparison of power and performance with traditional AI solutions.

Keywords: DNN, xDNN FPGA, DSA, Inference, SoC, FPGA

KEYPHRASE EXTRACTION FROM UNSTRUCTURED TEXT

Abhishek Pandey

Emerging Tech
Fidelity Investments
Bangalore, India

abhishek.pandey@fmr.com

Rohit Gadia

Emerging Tech
Fidelity Investments
Bangalore, India

rohit.gadia@fmr.com

Mridul Mishra

Emerging Tech
Fidelity Investments
Bangalore, India

mridul.mishra@fmr.com

Today, the cognitive load on analysts has increased exponentially as they have to process huge amount of information, mostly text, such as annual reports, research notes, news articles and earning calls transcript etc. We explored different methodologies (supervised and unsupervised) that can be used to extract topics/keywords/phrase present in unstructured text and assist analyst to quickly find the relevant information. In this paper, we have discussed our solution which helps find topics/keywords/themes from the unstructured text.

A COMPARATIVE STUDY OF BOOSTING ALGORITHMS FOR PATENT GRANT DURATION PREDICTION

Raman Dutt

Department of Computer Science
Shiv Nadar University
Uttar Pradesh, India
rd650@snu.edu.in

Dr. Vinita Krishna

School of Management and Entrepreneurship
Shiv Nadar University
Uttar Pradesh, India
vinita.krishna@snu.edu.in

The rapid pace of technological advancement coupled with the flood of data since the past decade has led the companies to put on their thinking hat in terms of aligning their business goals with competitive edge in the market. Moreover, in the wake of growing importance of intangibles as Intellectual Property (IP), patent data has been an obvious choice for analysis with their ease of accessibility, availability in large amounts and recent advancements in the field of Machine Learning. This work attempts to discuss the traditional approaches applied to patent data analysis and weigh the benefits & constraints of these approaches in patent- analysis and business intelligence.

In doing so, we identify some of the important factors-the so called patent characteristics, cited in literature as impacting the decision on grant duration of patent applications. We also show a comparative study between different prediction algorithms particularly belonging to the class of Boosting Algorithms.

Different algorithms ranging from those developed earlier in the day to some of the most recent ones have been used in our work. These algorithms are run in a hit and try mode to get the predictive models for patent grant duration. The results are then compared to check which algorithm offers the best solution. We see that the results obtained are in harmony with the general trends in patent grant duration in the real world. Our research is exploratory in nature and to the best of our knowledge, first of its kind in terms of country focus (India), the specific dataset (Indian Patent Office) and technique (ML/DL) used in patent grant duration prediction.

Keywords: *boosting, business intelligence, machine learning, patent analysis.*

DOUBLE SIEVE SELECTION MODEL FOR ABSTRACTIVE TEXT SUMMARIZATION

Lavanya Venkatasubramaniam

Computer Science Department
National Institute of Technology
Tiruchirappalli, India
vlavanya418@gmail.com

Kavitha Subramaniam

Computer Science Department
National Institute of Technology
Tiruchirappalli, India
kavithasubramaniam.nitt@gmail.com

S. Jaya Nirmala

Computer Science Department
National Institute of Technology
Tiruchirappalli, India
sjaya@nitt.edu

The Double Sieve Model for Abstractive Text Summarization introduces a two-layer selector function (Sieves) to the Attentional Encoder-Decoder Recurrent Neural Network architecture. The model consists of a bidirectional GRU encoder, the Double Sieve Selectors and a unidirectional GRU decoder. The sieves help to generate the best possible words for an effective summary by enhancing the contextual understanding and eliminating less important words in a two-step integrated selection process bridging the encoder and decoder. Other features such as keywords modelling and a beam search for the decoder are also added. The model is evaluated on the English Gigaword dataset. The experimental results show that the proposed model architecture produces more accurate, perceptive and cohesive abstractive summaries.

Keywords: *Abstractive Text Summarization, Recurrent Neural Networks, Natural Language Processing.*

BRAIN IMAGE SEGMENTATION FOR TUMOR DETECTION USING DEEP LEARNING

V. Gunamathi, M. Rashmi, H. Nilofar Nisa
Department of IT

K.V. Uma
Assistant Professor,
Department of IT
Thiagarajar College of Engineering,
Madurai, India
kvuit@tce.edu

Brain tumor is unwanted growth of unhealthy cell which increase intracranial pressure within skull. Segmenting brain images could be highly beneficial for diagnosing, treating and evaluating the progress of specific diseases. Up to this point, manual segmentation, performed by experts, is the conventional method in hospitals and clinical environments. Although manual segmentation is accurate, it is time consuming, expensive and might not be reliable. Nowadays, Brain tumor segmentation and detection of tumor is one of the most challenging and time-consuming tasks in medical image processing. Magnetic resonance imaging (MRI) can be used to create an image of any part in a body and it provides an efficient and fast way for diagnosis of the brain tumor. In this paper, we present a fully automatic brain tumor segmentation method based on convolution neural networks (CNN) . For accurate classification of tumor versus non-tumor brain CNN is used. Firstly, the MRI image of the patient's brain is collected and fed to the program which is then compared with the humungous number of brain MRI-s in the database from existing. Different activation function and number of epochs are changed and the performance is evaluated.

Keywords: *Magnetic Resonance Imaging, Machine Learning, Convolution Neural Network, Segmentation and Detection.*

COMPROMISED MACHINE LEARNING SHIELD – NEW ERA OF SECURITY EPIDEMIC

Ashima Purohit
Paypal India Pvt. Ltd.
ashimaj@gmail.com

Artificial Intelligence (AI) – and Machine Learning (ML), with its ability to predict based on previously seen data, has become an integral part of modern-day digital applications. Machine Learning models are finding an essential place in virtually every industry, be it medicine, finance, entertainment, law enforcement or cybersecurity, name it and machine learning is there.

Since Machine learning is increasingly being used at the core of several critical applications, such as for self-driving cars, drug recommendation systems, high-volume trading algorithms, privacy and security of sensitive data etc.; any adversarial manipulation on model, can lead to devastating results.

Wondering what would it be like to have your machine learning (ML) model come under security attack, yes COMPROMISED? Have you thought through how to check/monitor security attacks on your AI/ML models?

Historically less attention has been paid to the ways in which AI can be used maliciously. ML models, much like any piece of software, are prone to theft and subsequent reverse engineering. Machine learning is susceptible to adversarial activity, where an attacker can manipulate the input data to deceive the deployed ML model.

This paper aims to describe the potential threats associated with current methods of collecting or building ML systems and elaborate on the techniques to protect these models. The intention is to bridge the gap between machine learning and privacy and security technologies by helping attendees get acquainted with machine learning, the potential threats to privacy, the proposed solutions, and the challenges that lie ahead.

AUTOMATIC DETECTION OF UNKNOWN PERSONS TO ALERT KNOWN AUTHORITIES

Narayanan Durai

Unisys India Private Ltd,

Bangalore, India

Narayanan.Durai@Unisys.com

Santosh Kumar Gupta

Unisys India Private Ltd,

Bangalore, India

Santosh.Gupta@in.unisys.com

In today's world, it is important to automatically detect un-authorized, suspicious people from the video feed at a particular location, such as university campuses, office buildings, etc. Today's systems only allows us to react to crimes. Using AI and ML in facial recognition, we can proactively identify unknown persons and send notification to the authorities such as security persons, police, etc. to address that matter before any potential offense or criminal activity can occur. It involves two phases such as Registration to register the known persons and Recognition to recognize the persons either as known or unknown persons by comparing the known persons identified in Registration phase. Recognition phase involves face detection, face cluster, save face features and face recognition using machine-learning algorithms.

Keywords: *Face Detection, Face Cluster, Face Recognition, Save Face Features.*

A COMPARISON STUDY ON REGRESSION TECHNIQUES IN HOUSE PRICE PREDICTION

Manasa J

Research Scholar, Mathematics
Visveshwariah Technological University
Bengaluru, India
mansa.chandan@gmail.com

Narahari N S

Department of IEM.
RV College of Engineering
Bengaluru, India
naraharins@rvce.edu.in

Radha Gupta

Department of Mathematics
Dayananada Sagar College of Engineering
Bengaluru, India
radha.gaurav.gupta@gmail.com

Prediction models for determining the sale price of houses in cities like Bengaluru is a challenging task. The sale price of properties in cities like Bengaluru depends on a number of interdependent factors. The key factors that influence the price includes area of the property, location of the property and also facilities and amenities if it is a residential property like a house, number of bedrooms, size of the apartment, power generation facilities and many other such factors. This paper is based on an analytical study by considering the publicly available data set pertaining to housing property available in machine hackathon platform.

The data has nine features. The attempt in this work is to build a prediction model for determining the price based on factors influencing the price. The explorations of modelling deploy some of the regression techniques like the multiple linear regression (Least Squares), Lasso and Ridge regression models.

These models are employed in building the predictive model and a comparison study of the errors in prediction among these models have been made to select the best performing model.

Keywords: *House price prediction, Lasso regression, Ridge Regression, Multiple linear regression*

CUSTOMER VOICE TO SOLVE OUT OF STOCK

Raji Nair

Data Analytics and Computational Intelligence (DACI)
Lowe's Services India Pvt Ltd.
Bangalore, India
raji.nair@lowes.com

Sanhita Mukherjee

Data Analytics and Computational Intelligence (DACI)
Lowe's Services India Pvt Ltd.
Bangalore, India
sanhita.mukherjee@lowes.com

V Shekar

Data Analytics and Computational Intelligence (DACI)
Lowe's Services India Pvt Ltd.
Bangalore, India
v.shekar@lowes.com

“Out of stock” is one of the top reasons for customer (dis)satisfaction, among traditional retailers. Retailers are missing out on nearly \$1 trillion in sales because they don't have on hand what customers want to buy in their stores, according to a study of about 600 households and several retailers conducted by IHL Group.

Previous studies also show that the average Out of Stock rate is about 8%. That means that one out of 13 products are not purchasable in the exact moment the customer wants to get it in the store. This rate rises to 10% or even more when there are promotions or discounted products.

This issue is troubling for retailers, considering that their customers are already turning online, in many cases to Amazon, to find things they can't or won't look for in stores. Retailers themselves are exacerbating the problem as they boost omnichannel services like click-and-collect, which deplete shelves further and take up staff time.

This paper tries to comprehend this specific retail problem from a customer lens by deep diving into customer voice data and trying to identify specific product class that was “out of stock” in different regions leading to customer dissatisfaction. Loyalty surveys are conducted regularly, across different touchpoints with customers to understand their shopping behaviour and perception towards the company/brand. Through In-store purchase experience survey, millions of responses are captured on an annual basis from the customers on their shopping experience. These surveys can also record the specific products that the customers found Out of Stock. The objective was to classify these comments into specific Product Class which can be shared with the store managers and Merchandizing divisions. Without this analysis, it would have to be read manually to quantify the “out of stock” products.

A two-pronged approach of “Bag of words” and “Supervised Learning” was used to solve this business problem and improve the accuracy of the classification. This process is explained in detail in the paper.

Feedback Data: Customer Voice

Keywords: *Text Analytics, Data Mining, Supervised Learning, Out of Stock.*

SENTIMENT ANALYSIS ON TWITTER HASHTAGS –A STUDY ON OPINION AND EMOTIONS

Lokesh B

PSG Institute of Management
PSG College of Technology
Coimbatore, Tamil Nadu

lokeshbalasundaram@gmail.com

Dr A. Mansurali

Assistant Professor (Senior Grade)
PSG Institute of Management
PSG College of Technology
Coimbatore, Tamil Nadu

mansurali@psgim.ac.in

Harshini J

PSG Institute of Management
PSG College of Technology
Coimbatore, Tamil Nadu

jharshini27@gmail.com

Sentiment analysis about twitter tweets & hashtags employs various text mining techniques allowing companies to analyze the posts and associated sentiments with the tweets to find key topics or themes which resonate well with audience. Even the leaders of the country like Presidents & Prime Ministers of nations and various international organizations like WTO, UN, etc. are sending out its messages to public via tweets. These tweets contains Hashtags starting with “#” followed by keywords or phrases to express opinions about common point of interest. In this paper we perform sentiment analysis, to identify and extract opinion of people thoughts on hashtag “#Pray_for_Nesamani” from the tweets. This hashtag has become a worldwide trending topic of discussion pushing down the swearing-in ceremony of Indian prime minister and this hashtag was used by different business entities like Air Asia, Nippon paints, Xiaomi etc. and various state government departments like Tamil Nadu police, Ahmedabad police etc. to promote their brands and social messages since it became a worldwide trending topic of discussion. We have used R programming language to extract, clean the tweets from twitter and analyzed the tweets to find 1) Sentiment-bearing Phrase or keyword, 2) hashtag level sentiment polarity and emotions behind the tweets 3) sentiment score of each phrase and keyword.

Keywords: *Sentiment analysis, Opinion mining, Text mining, Twitter, Hashtag, Emotions.*

PLATFORM EVALUATION FOR CHATBOTS: WHICH ONE FITS THE BUSINESS NEEDS?

Ruma Mukherjee

Unisys Corporation

Bangalore, India

ruma.mukherjee@unisys.com

Kamisetty Yamini Preethi

Unisys Corporation

Bangalore, India

kamisetty.preethi@unisys.com

In the evolving world of Artificial Intelligence (AI), Chatbots are forging ahead in the conversational service space.

Chatbots help in accelerating the businesses, by automating complex tasks using intelligent interactive skills. In recent days, AI powered Chatbots are replacing rule based conversational agents. The Chatbots are expected to behave in an intelligent fashion with features like automated learning embedded in them.

Though there exists a large number of Chatbot platforms in the market, only a handful of them suit the contextual requirements for a specific business. In this paper, we detail criteria to select the appropriate platform for a specific use case. We use the criteria to compare four major platforms Google Dialogflow, Amazon Lex, Microsoft Azure Bot Service and Rasa to determine which fits the most business needs. The platforms are evaluated on a wide range of criteria including Natural Language Processing functionalities, automatic learning capabilities, ability to integrate custom models, methods of hosting, loose coupling with external messaging platforms, ease of deployment, scalability, security and other built-in features. This work provides a reference implementation of Chatbot for each platform, which the businesses can use as a reference for understanding the functionalities of each platform. The results of the exploration are intended for anyone who needs to build an intelligent Chatbot for a specific use.

Keywords: *Artificial Intelligence, Natural Language Processing, Platform evaluation, Automated learning.*

THE POWER OF ENSEMBLE LEARNING APPLIED FOR FRAUD DETECTION AND SALES CYCLE IMPROVEMENTS

Vinod Alangaram

Business Intelligence & Data Analytics

Fidelity Investments

Chennai, India

vinodalangaram@gmail.com

Ramanathan Natarajan

Business Intelligence & Data Analytics

Fidelity Investments

Chennai, India

Ramanathan.n2002@gmail.com

As we all know, the Objective of Machine Learning is to build a mathematical model that can accurately predict the outcome. Ensemble Learning is an advanced technique to improve the accuracy of prediction/classification outcomes by just not relying on a single model but instead making use of the Decisions from various applicable models. Ensemble systems are essentially Multiple Classifier Systems comprising of diverse models that factor in a confidence to the decisions made by each model, and even help in selecting optimal features for a better model. This is all with the understanding that the combined nature of models could potentially error-correct within themselves and eventually aid in improving the overall accuracy of the system. However, Diversity is the key as the intent is that the errors from each of the models when combined strategically could reduce the total error and improve performance & reliability.

Besides these, Ensemble models could be extremely useful in both cases when there are large amounts of data in consideration or when there is a lack of adequate data. Ensemble systems use built in Combination rules as in the case of Bagging, Boosting, AdaBoost and Stacking or classifiers can then be custom combined using one of several different combination rules.

Based on our experience we believe Ensemble learning techniques improve reliability on Machine learning systems and are essentially versatile, powerful, easy to apply and effectively benefit yielding across most of machine learning problems that cut across domains such as:

- Improve the performance of a Classification model that detects potential Fraudulent transactions
- Enhance or positively influence the sales teams to optimize and improve their Sales cycles

The objective of this paper is to introduce, explore and describe the advantages of Ensemble Learning techniques and to highlight the broad outline of Machine Learning problem scenarios where this could be potentially applied and eventually demonstrate the value in terms of performance improvements that this technique offers in those cases.

PREDICTION MODELLING FOR CUSTOMERS UPSELLING – APPLICATION OF VARIOUS MACHINE LEARNING ALGORITHMS

Dr. A. Mansurali

Assistant Professor (Senior Grade)
PSG Institute of Management
PSG College of Technology
Coimbatore, Tamil Nadu
mansurali@psgim.ac.in

V. Harish

Assistant Professor
PSG Institute of Management
PSG College of Technology
Coimbatore, Tamil Nadu
harish@psgim.ac.in

Lokesh B

PSG Institute of Management
PSG College of Technology
Coimbatore, Tamil Nadu
lokeshbalasundaram@gmail.com

Banks with a customer base is always looking out to convert its existing customers for up selling their products. Selecting the existing customers for up selling the products is beneficial in many ways like targeting the known customers, targeting the right customers, cost of acquiring the customer is less and the behavior and profile of the customer is known. A banking dataset has been chosen for the purpose of applying various algorithms. Dataset contains 5000 data points, 12 predictor variables and one outcome variable which is whether the customer has reacted to the campaign for a personal loan or not. The success rate of the previous campaign was about 9 percent. The problem at the hand now is to identify existing clients that have higher chance to subscribe for purchasing a loan or for any new products and focus marketing effort on such clients. The objective of the study is to apply a classification approach to predict which clients are more likely to subscribe for personal loans using various models and machine learning algorithms and select the best algorithm based on the different evaluation parameters say AUC, Classification Matrix etc... This will increase the success ratio for up selling while at the same time reduce the cost of the campaign. The Various algorithms applied are logistic regression, CART, neural network etc..., using R. The output of the model is further used to predict the probable customers to target and also to understand their profile characteristics.

Keywords: *Bank, Predictor, Classification, Model, Marketing.*

MODEL SELECTION FOR OVER DISPERSED COUNT DATA USING CLASSIFICATION TECHNIQUES

K.M. Sakthivel

Department of Statistics,
Bharathiar University,
Coimbatore - 641046,
Tamil Nadu, India

C.S. Rajitha

Department of Mathematics,
Amrita School of Engineering,
Amrita Viswa Vidyapeetham,
Coimbatore, India

There are many fields in reality the data observed are of the nature over dispersed such as biological study, non-life insurance, accident analysis etc. The researchers have already used mixed models and zero inflated models for modeling this kind of data. Among these models, zero inflated Poisson (ZIP) models and hurdle models are the most commonly used models. In the era of machine learning, one can use Artificial Neural Network (ANN) for modeling this kind of data. In this paper, we compared the performance of various models such as ANN, ZIP and hurdle models in terms of mean square error, standard error, bias and relative efficiency using a motorcycle insurance data set. And we also studied suitability of these models using classification techniques such as discriminant analysis, CART and random forest for modeling over dispersed count data. Further we proposed an algorithm for selecting a better model by comparing misclassification rates using different classifiers.

Keywords: *Artificial Neural Network, zero inflated models, discriminant analysis, CART, random forest.*

DETECTION OF THYROID DISEASE BY USING FEATURE SELECTION AND CLASSIFICATION

K.V. Uma

Assistant Professor

Department of Information Technology

Thiagarajar College of Engineering

Madurai, India

kvuit@tce.edu

P. Shobana

Department of Information Technology

Thiagarajar College of Engineering

Madurai, India

pshobanashalini18@gmail.com

S. Nandhini

Department of Information Technology

Thiagarajar College of Engineering

Madurai, India

nandhinisel19@gmail.com

J. Keerthi

Department of Information Technology

Thiagarajar College of Engineering

Madurai, India

jkeerthimeena6@gmail.com

Thyroid gland produces thyroid hormones to help the regulation of the body's metabolism. The abnormalities of producing thyroid hormones are divided into two categories. Hypothyroidism which is related to production of insufficient thyroid hormone and hyperthyroidism related to production of excessive thyroid hormone. Separating these two diseases is very important for thyroid diseases. In this paper we use Joint Mutual Information (JMI), Minimum-Redundancy-Maximum-Relevance-Selection (MRMR) and Conditional Mutual Information Maximization (CMIM) Algorithms to find the best features in Thyroid diseases identification. The selected features are given as the input to the classifiers such as NBC (Naïve Bayes Classifier), SVM (Support Vector Machine), C5.0, Random Forest, and KNN (K Nearest Neighbours). The performance of these classifiers are compared together in order to select the best classifier for predicting the Thyroid Disease. Results are validated using Hypothyroid dataset taken from Kaggle.

Keywords: *Hypothyroidism, Hyperthyroidism, JMI, MRMR, CMI, NBC, SVM, C5.0, Random Forest, KNN.*

OPTIMIZATION OF LOGISTIC REGRESSION USING ELEPHANT HERD OPTIMIZATION

Abijith Pradeep, Aswin Kumar K S, Dileep Vijayakumar, Paul Antony

Dept. of Computer Science and Engineering
Federal Institute of Science And Technology (FISAT)
Kochi, India

Pankaj Kumar G

Asst. Professor
Dept. of Computer Science and Engineering
Federal Institute of Science And Technology (FISAT)
Kochi, India
pankaj.ekm@gmail.com

This paper introduces a method of optimizing logistic regression by utilizing Elephant Herd Optimization (EHO).

The optimized logistic regression is then used to predict breast cancer. The proposed method proved to be more efficient compared to the existing methods.

Keywords: *Elephant Herding Optimization, Gradient descent, Logistic regression, Breast cancer detection.*



*Analytics
Society of India*

Thank You