



National Conference
on
Machine Learning and Artificial Intelligence
30-31 October, 2023

Conference Proceedings

Organized by
CET School of Management
College of Engineering Trivandrum

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

BUILDING ROBUST PROCESSES THROUGH APPLICATION OF LOW CODE ML & AI MODELS

Anurag Seksaria
Managing Partner
Anurag Seksaria Management Consulting
Bangalore, India
hello@anuragseksaria.org

The authors will like to present an approach to build robust processes using Low Code Machine Learning and Artificial Intelligence Models specifically in continuous industry like Chemicals, Cement, Metals where major component of raw material is natural resource like Iron Ore, Bauxite, Coal, Lime where the quality parameters have high and uncontrollable variations.

The authors will like to present an approach to build robust processes using Low Code Machine Learning and Artificial Intelligence Models specifically in continuous industry like Chemicals, Cement, Metals where major component of raw material is natural resource like Iron Ore, Bauxite, Coal, Lime where the quality parameters have high and uncontrollable variations.

Like any other industry, the quality parameters like Purity, Yield, Moisture Content are function of Raw Material Quality (X_{RM}), Process Parameters (X_{PP}) and Intermittent Product Parameters (X_{IP})

$$Y \text{ (Output Parameter)} = f \{ X_{RM1}, X_{RM2}, \dots, X_{PP1}, X_{PP2}, \dots, X_{IP1}, X_{IP2}, \dots \} + e$$

Author has tried the popular UOSEMN model to create predicable models and deploy them in manufacturing processes to enhance Quality and Productivity with minimal human interventions.

Understand Business Requirement: While it appears obvious, its of utmost importance to understand the expectation of the business stakeholders.

Obtain Data: Understand and compile data for end-to-end process, starting from Raw Material, Inprocess Checks and Finals validation of the products.

Scrub: Although both quality of and quantum of the data available has improved many folds, a careful scrubbing of data is essential to getting predictable models. Any short cut during this stage may cost multiple iterations and delay in getting the desired results. Continuous process industry has additional constraints in terms of

- a. Time Lag between the processes
- b. Heterogeneity in the raw material, in process and final product
- c. Collinearity, that is relationships between the influencing factors.

Model: Build Predictable Machine Learning Models using methods like Logistics Regression & Decision Tree. No Low or No Code approach helps in engaging domain experts in building their own models which can be generalized but remain flexible.

iNterpret : The business insights are drawn and processes are optimized with close collaboration of management teams. Post the pilot runs, AI algorithms are created to deploy the model and accompolish in process adjustments.

Author experiences large scale deployments resulting in transformational improvement in Efficiency, Effectiveness and Stakeholder Experience.

Keywords: Low Code Machine Learning Models, Time Lag, Heterogeneity, Collinearity, Decision Tree, Logistics Regression

EXPLORING THE FACTORS THAT ENABLE AI APPLICATION IN FINANCIAL FRUAD

Gouri Anilkumar

Student

Thiruvananthapuram, India

anilkumargouri750@gmail.com

Dr. Suresh Subramoniam

Director

CET School of Management

Thiruvananthapuram, India

sureshsubramoniam@gmail.com

Financial institutions and governments throughout the world have been quiet concerned about financial crime. The most prevalent financial crimes that endanger the stability of financial systems are fraud, money laundering and cybercrime. Artificial intelligence and machine learning algorithms are used in the rapidly expanding field of AI-enabled financial crime detection to spot and stop financial crimes. Financial institutions can enhance their compliance efforts and lower the risk of financial crime by adopting AI in financial crime detection to identify suspicious trends and unusual patterns. In this exploratory study attempt has been made to-identify some of the major factors that enable AI to spot and prevent potential financial crimes where traditional methods fail. The items were identified from literature, and an exploratory factor analysis was conducted. A total of 20 items were identified and a questionnaire based on these items were formulated and responses were collected online. After the analysis the 20 items were grouped under five factors and these factors were termed as technology, accuracy, integration, adaptability and speed.

Keywords: Artificial intelligence, Financial crimes, Detection, Identified

ENHANCING TAXPAYER RISK PREDICTION THROUGH LLM-DRIVEN PROFILE TUNING

Shubhradeep Nandi

Sr. Data Scientist

Data Analytics Unit

GoAP Vijayawada, India

shubnandi.ai@gmail.com

Kalpita Roy

Sr Data Scientist

KSA Bengaluru, India

kalpita.roy99@gmail.com

Taxpayer risk prediction is pivotal for governmental financial departments globally. While traditional methodologies have automated the detection of potential risky taxpayers, their efficacy is waning due to two primary impediments: the continued reliance on outdated algorithms amidst the rapid emergence of generative AI and Large Language Models (LLMs), and the absence of a standardized metric to delineate genuine taxpayers. Addressing these challenges, this research introduces a groundbreaking approach centred on LLMs for taxpayer risk assessment. By integrating taxpayer transactional data into predefined templates, we harness LLMs to generate comprehensive natural-language profiles. These profiles are then employed to fine-tune expansive pre-trained models, facilitating nuanced risk predictions. Preliminary evaluations against established baselines attest to the superior performance of our proposed methodology. This research not only underscores the transformative potential of LLMs in financial oversight but also paves the way for future explorations into AI-driven fiscal governance.

Keywords: Taxpayer Risk Prediction, Large Language Models (LLMs), Financial Oversight, AI-Driven Fiscal Governance, Natural-Language Profiles.

ONDC ADOPTION BY SELLERS: KEY CHALLENGES

Ganga Susan Kurian

MBA student
CET School of Management
Trivandrum, India
ganga_susan@yahoo.co.in

Dr. Mithra Manmadhan

Assistant Professor
CET School of Management
Trivandrum, India
mithramohan18@gmail.com

Dr. Nima Ravi

Assistant Professor
CET School of Management
Trivandrum, India
nimaravi@gmail.com

India has made significant progress in its digital transformation journey creating an impressive global footprint with its futuristic digital innovations, supporting technologies and scalable architecture. India's digital landscape got a favorable environment for rapid growth because of increased smartphone adoption by the general public, cost-effective high-speed data plans, user-friendly digital payment channels, a large IT talent pool and supportive policies by the government for digitalization. India's flagship program 'Digital India' along with 'India stack' are the building blocks that accelerated the development of Digital Public Infrastructure (DPI). DPI initiatives began with the creation of a foundation of unique identity layer of Aadhar, complementary payment systems like UPI followed by data layer.

Digital commerce has a high potential to fuel economic growth further but current penetration is only 7 percent of overall market restricted by multiple barriers. In 2021, India kickstarted its ambitious plan to democratize e-commerce with ONDC (Open Network for Digital Commerce) envisioned to create an inclusive ecosystem for e-commerce that enables even the smallest of Indian merchants to reap benefits of e-commerce. Currently, Indian ecommerce is dominated by large players which is out of reach for the majority of MSMEs and rural players. ONDC program is intended to curb digital monopolies/duopolies and make it a level playing field with higher inclusion. It aspires to become UPI of e-commerce by making e-commerce procedures open source, non-platform-based, decoupled, inter-operable, transparent, scalable, secure and cost effective.

ONDC was incorporated in 2021 as a Section 8 company with dedicated advisory council. Key roles in ONDC include buyer and seller network participants, logistic service providers and ecosystem participants. Pilot phase was launched in 2022 in grocery and food delivery segments in five major cities. The current statistics shows an increased presence in 273 cities with 50K plus registered sellers but the growth is much lesser compared to 4 lakh plus sellers in major e-commerce platforms. But there are quite a number of challenges ONDC needs to overcome. Implementation of ONDC is much complex than that of UPI considering the number

of players and steps involved. Other challenges include high competition from established players, ensuring product quality, user awareness and engagement, timely shipments, after-sales support (returns and refunds), technology support and data privacy. It is important for the government to understand the most pressing challenges faced by sellers in adopting ONDC, so that changes may be brought in or support may be given by the government/ONDC. This study strives to identify the key challenges faced in the adoption of ONDC by sellers using Fuzzy AHP approach. The results of the study can significantly aid in the successful implementation and widespread acceptance of this digital commerce framework by ensuring that the platform is responsive to the needs of its users . Thus ONDC can position itself as a viable and attractive option for sellers in the Indian market which in turn can lead to fostering a larger and more vibrant digital commerce ecosystem becoming another example for the world similar to the UPI program.

Keywords: ONDC, Open Network for Digital Commerce, Challenges India stack, DPI

FACIAL ASYMMETRY: A BEHAVIOMETRIC INDEX FOR ASSESSMENT DURING A FACE-TO-FACE INTERVIEW

Dr. Tanusree Dutta

Associate Professor
Indian Institute of Management Ranchi
Ranchi, India
tanusree@iimranchi.ac.in

Shuvam Keshari

Student
Indian Institute of Technology Kharagpur
Kharagpur, India
skeshari1998@gmail.com

Raju Mullick

Research Scholar
Indian Institute of Technology Kharagpur
Kharagpur, India
raju.mullick@gmail.com

Ashish Rathor

Student
Indian Institute of Technology Kharagpur
Kharagpur, India
ashishrathor72@gmail.com

Prof. Priyadarshi Patnaik

Professor, Department of Humanities & Social Sciences
Indian Institute of Technology Kharagpur
Kharagpur, India
priyadarshi1@yahoo.com

Choosing the right person for the right job makes the personnel interview process a cognitively demanding task. Psychometric and biometric tests have often been used to aid the process, but they have their own limitations. We propose the use of behavioumetry as an assistive tool to facilitate an objective assessment of the interviewee without increasing the cognitive load of the interviewer. Behavioumetry analysis thin slices of behavior and provide unbiased information about the interviewee. The current study details the methodology of the behavioumetry tool to capture asymmetry and micro expressions of the face. Hemi-facial composites using a structural similarity index were used to develop a progressive time graph of facial asymmetry. In the pilot test, frame by frame analysis was performed on 3 youtube video samples, where Structural similarity index (SSID) scores of 75% and more showed behavioral congruence.

Keywords: behavioumetry, facial asymmetry, micro expressions, hemi-facial, congruence

A MULTI-PRONGED APPROACH TO PREDICT REPRODUCTION NUMBER (R_0) OF SARS-CoV2 TRANSMISSION IN INDIA

Mrinmoy Roy

Associate Professor
PES University
Bangalore, India
mroy2612@gmail.com

Gopal Krishna Patra

Chief Scientist
CSIR – 4PI
Bangalore, India
gkpatra@csir4pi.in

The sudden rapid outbreak of the coronavirus has posed significant challenges to all communities. Researchers and doctors have struggled to fully comprehend all aspects associated with SARS-CoV2, such as its transmission rate, due to the virus's unique nature and the emergence of new strains. Making it incredibly tough to simulate accurately. Even as more data and information become available, these characteristics remain dynamic, making it impossible to quantify them. The lack of good models and sufficient data on confirmed cases and deaths at the start of the pandemic made it difficult to anticipate future cases and plan for key hospital resource needs and effective mitigation efforts. Because of the coronavirus's unfamiliarity and novelty, contradictory messages were transmitted about the outbreak's potential severity, the importance of mask-wearing and social distancing, and other related precautions, resulting in ad hoc, unplanned lockdowns and a rise in SARS-CoV2 cases. Supply chains and hospitals around the world were unable to keep up with the demand for ventilators, personal protective equipment (PPE), and intensive care unit (ICU) beds as doctors and scientists began to better understand SARS-CoV2's destructive and disruptive nature, causing critically ill patients to be turned away. The world's strongest and largest economies came to their knees, and resources have been stressed out due to the speed of coverage of the pandemic and the rate of transmission of the disease. Seeing the magnitude of growth in a number of cases also its consequences on the health administration of the country, an accurate prediction seems to be a very important requirement to control this in future and be prepared for the same. In this research, Mathematical, Statistical and LSTM modelling and identification techniques are applied in developing a prediction system to forecast the spread of SARS-CoV2 in India. When any Epidemic or Pandemic breaks out there is a need for immediate measures in exponential cases as it's evident that developing a vaccine would take some time, so nontherapeutic interventions such as social distancing or contact tracing have benefitted to delay the rising of SARS-CoV2 cases. Through this research, we are proposing a multi-pronged approach for bracketing the SARS-CoV2 cases in India. We found ARIMA model grossly underestimates, SEIRD model grossly overestimates, the Polynomial has shown promises for a long-term prediction (~20 days), The changing degree of polynomials for best fit, Deep learning models seems to have immense potential to bracket the number of cases especially for short-term (~10 days). A Comprehensive platform (most likely a hybrid one with DL, Statistical and Mathematical) is very useful to Sharpen the bracket as much as possible and use of clustering algorithms to identify vulnerable districts to potentially become SARS-CoV2 hotspots. Estimate changing R_0 (basic reproduction number) from data using Deep Learning, influenced by lockdown and un-lockdown patterns and design downscaling approaches.

Anaconda environment with Python, Jupyter Notebook and Google Collab ML platforms are used as simulation tools for the algorithms to test and predict Indian SARS-CoV2 disease data.

Keywords: Multipronged Approach, SARS-CoV2, ARIMA, SEIRD, Deep Learning

SUMMARISING PATIENT-PHYSICIAN CALLS USING FINE-TUNED BART MODEL

Charankumar P C

Student

R.M.K Engineering College

Chennai, India

char20110.ad@rmkec.ac.in

Sumanth Vamsi M

Student

R.M.K Engineering College

Chennai, India

suma20217.ad@rmkec.ac.in

Sunil B

Student

R.M.K Engineering College

Chennai, India

suni20218.ad@rmkec.ac.in

Nishant Y P

Student

R.M.K Engineering College

Chennai, India

nish20227.ad@rmkec.ac.in

Documents containing biomedical information come in a variety of formats. In healthcare industry, physicians often have phone or online conversations with patients to address medical concerns, offer guidance, and respond to inquiries. Summarizing this information are very crucial but, manual summarization of transcripts result in challenges such as the task of manual consumes significant time, complexity of the task increases exponentially and the probability of errors rises considerably, making it more prone to inaccuracies. Pretrained language models have served as important backbones for natural language processing. To ensure optimal usability and efficiency, users are provided with concise summaries that capture key details and treatment plans discussed in their conversations by integrating the finetuned bidirectional autoregressive transformer (BART) model trained using GPT-3 synthetic dataset, renowned for its impressive summarization. Additionally, artificial intelligence (AI) analysis for the obtained summaries is performed. The analysis contains the satisfaction score, key words spoken in the conversation, key sentences in the conversation and also topic modelling that provides a lifestyle plans for the users. Performance metrics such as ROUGE-1, ROUGE-2, ROUGE-L indicate that bart-large-cnn outperform other state-of-the-art models such as bart-base and t5-large. Incorporating FRAMEVR, we venture into immersive VR development. The framework facilitates a comprehensive environment, enriching user engagement. For accurate and high-quality transcription of audio and video files, we employ the state-of-the-art Whisper AI model, which excels in the Speech-to-Text task and provides multilingual support. Additionally, cloud synchronization heightens data integrity and thwarts unsanctioned alterations of stored transcripts and summaries.

Keywords: BART, ROUGE, FRAMEVR, Whisper AI, Cloud

RETAIL PRICE OPTIMISATION THROUGH MULTI-ARMED BANDIT EXPERIMENTS

V Shekar

Sr. Data Scientist, PPCI

Lowe's Services India Pvt Ltd

Bengaluru, India

v.shekar@lowes.com

Optimising the prices of products in the retail landscape while carrying a wide range of products remains a challenge. The breadth of the product catalogue, varying demand behaviours across categories, stock availability, competitiveness, changes in macroenvironment, changes in demand patterns, or even the absence of demand patterns elevate the complexity of pricing products. A near-optimal price will always reap benefits for both retailers and their customers. Selling items at deep discounts would negatively impact profit margins, and pricing items higher would impact revenue and customer satisfaction. In order to successfully find an optimal price to balance both profitability and revenue, price optimisation is vital. Often, optimising the price of a single item would require business knowledge, understanding its historical price sensitivity, customers willingness to pay, and many other factors. Due to this, many retailers price items manually. But very few test the price to observe sensitivity or elasticity to arrive at a pricing decision. Test-learn based price decisions have proven to identify optimal prices when done in real time. Frameworks like Multi-Armed Bandit Experiments help in testing multiple price points simultaneously and learning from the demand response through a finite number of actions. Also, with innovation in competitive intelligence and the ability for online retail channels to more quickly see and respond to demand responses, the useful life of an optimal price for the key value items in an industry is much shorter and sometimes an optimal price may need to be determined intraday. Therefore, use of this reinforcement learning framework at scale, the simplicity of real-time implementation to identify the current demand pattern for an item, helping in deciding the optimal price, are explained in this paper.

Keywords: Price Optimisation, Price Testing, Home Improvement Retail, Multi-Armed Bandit Experiments, Reinforcement Learning

SOCIAL CAUSE-DATA-DRIVEN APPROACH FOR DROUGHTEARLY WARNING USING SIESTASCALABLE PATTERN ANALYSIS

Christina Sharanyaa Selvakumari S

Department of Computer Science and Engineering
Panimalar Engineering College Chennai City Campus Chennai
Chennai, India
christysharankumari2005@gmail.com

Padmapriya J

Department of Computer Science and Engineering
Panimalar Engineering College Chennai City Campus Chennai
Chennai, India
padmapriyaj.panimalar@gmail.com

Murugvalli S

Department of Computer Science and Engineering
Panimalar Engineering College Chennai City Campus Chennai
Chennai, India
principal@panimalarengineeringcollegechennai.ac.in

Droughts are significant natural disasters which cost the life of humans and all living organisms. Hence this disaster require early detection and mitigation measures. The integration of big data analytics and DataOps methodologies has revolutionized drought prediction and mitigation strategies. This study explores an innovative approach to drought prediction and mitigation using pre-recorded data from multiple sources, utilizing big data analytics and DataOps methodologies. The approach involves data collection from satellite imagery, meteorological stations, hydrological databases, soil measurements, vegetation monitoring, and social-economic statistics. Preprocessing and feature engineering techniques are applied to ensure data quality and extract relevant indicators for drought prediction. Big data techniques, such as predictive modeling, DataOps, compression, and indexing, are used to develop predictive models that forecast drought conditions based on historical data and real-time inputs. The combined model, incorporating local, regional, and global data, achieves the best performance within acceptable timeframes. The study aims to develop an integrated approach for drought mitigation using data compression, SIESTA indexing, and DataOps practices to enhance the efficiency and effectiveness of drought prevention strategies. Data compression techniques handle massive volumes of log entries generated by data-intensive systems, while SIESTA indices support ad-hoc pattern queries in drought early-warning systems. DataOps principles ensure the sustainability of data-intensive applications, facilitating real-time monitoring and decisionmaking for effective drought prevention strategies. To conduct research on data reduction, DataOps, and drought prevention using big data, consider using publicly available datasets from reputable sources. These include Fusion XGC Dataset for scientific discovery, Climate Modeling Dataset for climate modeling, Log Entry Datasets for societal goals, World Hunger Data for world hunger, and AI-driven Early-Warning System for drought prevention. Acquire high-fidelity first-principle simulation data for fusion devices, climate modeling data, and world hunger data from relevant sources. DataOps can also be used for addressing societal goals, such as logging systems and world hunger data. Gather relevant datasets related to world hunger, agricultural production, food security, and social indicators to support the argument that DataOps can aid in addressing societal goals. Access to log data from organizations, IT companies, and government agencies is essential. A well-sourced and

relevant dataset is crucial for the validity and credibility of research findings in data reduction, DataOps, and AI-driven drought prevention. Data compression techniques, such as ZFP and SZ, can significantly reduce data sizes in scientific simulations, making large-scale fusion and climate modeling more feasible within resource constraints. The combination of data compression, SIESTA infrastructure, and DataOps methodologies enhances the efficiency and effectiveness of drought mitigation strategies, contributing to sustainable solutions for societal challenges like drought prevention. The reduced model approach, such as DuoModel, transforms data to better match compressor design philosophies, addressing the limitations of existing lossy compression methods. SIESTA infrastructure employs inverted indices and advanced query processing, enabling efficient detection of arbitrary patterns in non-streaming log data. DataOps methodologies play a crucial role in addressing complex societal challenges like drought prevention, with the AI-driven early-warning system developed through DataOps facilitating resource allocation decisions and providing precise predictions for drought mitigation efforts. The integration of data compression, SIESTA, and DataOps offers several significant implications for drought mitigation, including enhanced data handling, improved early-warning systems, sustainable drought management, resource optimization, and better decision-making. By leveraging these technologies, stakeholders can make informed decisions and take proactive measures to combat drought, fostering a more sustainable future for communities and ecosystems. This research highlights the potential of DataOps in addressing societal challenges like drought mitigation and emphasizes the importance of incorporating diverse data sources by compressing data and using SIESTA in decision-making processes.

Data Reduction for Extreme-Scale Scientific Simulations:

1.1 Problem Statement: The increasing volumes of data generated by simulation-based scientific discovery require efficient data management and analysis to overcome computational and storage constraints.

1.2 SIESTA Approach: Adopt DataOps principles to preprocess and compress scientific data effectively. Transform data to align with compression techniques, enhancing compression ratios for improved storage and analysis efficiency.

1.3 Dimension Reduction with SIESTA: Utilize dimension reduction techniques, such as PCA, SVD, and Wavelet, to synthesize reduced models directly from analysis output. Evaluate the impact of different configurations on compression ratios and performance.

1.4 SIESTA Contributions: Extend the SIESTA approach by exploring various inverted index forms and optimizations during query phases to further enhance response times and scalability.

Ad-Hoc Pattern Queries on Large-Scale Log Data:

2.1 Problem Statement: Organizations face challenges in efficiently querying large-scale log data for arbitrary patterns beyond traditional sequential pattern mining and complex event processing techniques.

2.2 SIESTA Approach: Apply SIESTA's DataOps principles to support ad-hoc pattern queries using inverted indices and a comprehensive query processor.

2.3 Advanced Query Processor: Enhance SIESTA's query processor to handle more complex queries over specific trace groups, expanding its capabilities for diverse applications.

2.4 Optimizations with SIESTA: Optimize preprocessing and querying phases with various configurations, leveraging the power of inverted indices and achieving faster response times in pattern detection queries. Evaluate scalability against competitors in Big Data scenarios.

2.5 Open-Source Implementation: Provide an open-source SIESTA implementation and extensive datasets for researchers to replicate and build upon.

Data-Driven Drought Prevention with AI-Driven Early-Warning Systems:

3.1 Problem Statement: Drought prevention requires efficient data-driven early-warning systems to address this wicked societal problem effectively.

3.2 SIESTA Approach: Employ SIESTA's DataOps principles to develop an AI software early-warning system for drought prevention, leveraging diverse data sources and AI models.

3.3 SIESTA Objectives: Determine resource acquisition decisions, design a project architecture applicable to other locations, and leverage big data sources for precise predictions within short timeframes.

3.4 SIESTA Findings: Identify key dimensions, such as precipitation, hydrological measurements, climate, and ocean oscillations, contributing to predictive power. Employ ensemble models for better interpretability and accuracy.

By combining data compression techniques, SIESTA's advanced infrastructure, and DataOps practices, this comprehensive approach aims to enhance decision-making, improve drought prediction accuracy, and support proactive measures to mitigate drought impacts. In conclusion, the combination of data compression, SIESTA, and DataOps offers a powerful toolkit for addressing the challenges of drought mitigation. These advanced technologies contribute to more sustainable and data-driven approaches, supporting key societal goals such as minimizing the impact of drought and enhancing agricultural resilience. By leveraging these technologies, stakeholders can make informed decisions and take proactive measures to combat drought, fostering a more sustainable future for communities and ecosystems alike.

Keywords: Big data, Lossy Compression, DataOps, SIESTA, Pattern Analysis

ROLE OF CHATBOTS IN THE AGRICULTURE SECTOR

Hemashree P1

Assistant Professor,
Coimbatore Institute of Technology
Coimbatore, TamilNadu
hemashree@cit.edu.in

Dhivya Darshini S S

PG Scholar
Coimbatore Institute of Technology
Coimbatore, TamilNadu
71762132014@cit.edu.in

Preethika T

PG Scholar
Coimbatore Institute of Technology
Coimbatore, TamilNadu
71762131037@cit.edu.in

The integration of chatbots in the agriculture sector has revolutionized the way farmers and stakeholders interact with technology. Chatbots, driven by artificial intelligence (AI), are intelligent virtual assistants capable of understanding and responding to human inquiries in realtime. They have emerged as powerful tools, providing valuable insights, guidance, and support to farmers throughout various stages of the agricultural value chain. This paper explores the role of chatbots in agriculture, highlighting their benefits, applications, and potential future developments. By analyzing real-world use cases and success stories, this paper aims to shed light on how chatbots can enhance productivity, efficiency, and sustainability in the agriculture sector. One of the primary benefits of chatbots in agriculture is their ability to deliver timely and accurate information. Farmers can interact with chatbots to obtain real-time weather updates, pest and disease management strategies, market prices, and personalized recommendations for crop cultivation. This instant access to information enables farmers to make informed decisions, optimize their farming practices, and mitigate risks. Chatbots also play a crucial role in providing advisory services. They can offer personalized guidance on crop selection, fertilizers, irrigation practices, and other farming techniques based on specific soil and weather conditions. Moreover, chatbots can assist farmers in diagnosing plant diseases and pests by analyzing images and providing suitable remedies. This not only empowers farmers with expert knowledge but also reduces reliance on physical consultations, making agricultural expertise more accessible to all. Furthermore, chatbots contribute to efficient supply chain management. They can assist in tracking and managing inventories, facilitating seamless communication between farmers, distributors, and retailers. By streamlining logistics and reducing wastage, chatbots help optimize resource allocation and promote sustainability in the agriculture sector. While the potential of chatbots in agriculture is immense, their successful implementation requires addressing challenges such as language barriers, connectivity issues in rural areas, and the need for continuous updates and improvements. Ethical considerations, such as data privacy and security, also need to be addressed to ensure user trust and adoption. Thus, the paper discusses about how chatbots have the potential to revolutionize the agriculture

sector by providing timely information, personalized guidance, and efficient supply chain management.

Keywords: Chatbots, Agriculture sector, Artificial Intelligence, Natural Language Processing(NLP), Effective communication

ENHANCING MARKETING STRATEGIES THROUGH RFM ANALYSIS AT YEKA HERBALS

Vidhya Lakshmi V

Research Assistant

Thiagarajar School of Management

Madurai, India

vidhyavijay73@gmail.com

Dr. Nataraj Balasubramanian

Assistant Professor

Thiagarajar School of Management

Madurai, India

bnataraj@tsm.ac.in

Purpose: This case study explores the utilization of RFM (Recency, Frequency, Monetary) analysis for customer segmentation at Yeka Herbals Pvt Ltd, an Indian sustainable beauty brand. The study's objective was to refine marketing strategies and bolster sales by discerning customer segments based on transaction behaviour. The research encompassed comprehensive data collection, meticulous preprocessing, rigorous analysis, and insightful interpretation, culminating in actionable findings tailored for precise targeted marketing campaigns.

Data Collection and Preprocessing: Transaction data spanning 2021 to 2022 were meticulously sourced from diverse platforms, including the company's website and various ecommerce portals. These data underwent meticulous cleansing, integration, transformation, and dimensionality reduction to ensure fidelity and consistency. Employing Power BI, the data were effectively visualized.

Data Analysis and Modelling: The foundation of the investigation entailed RFM analysis, which delved into Recency, Frequency, and Monetary value metrics. This exploration aimed to uncover the distribution of RFM scores, the trajectory of monetary value trends, and the creation of customer clusters. Employing a sophisticated two-step clustering algorithm combining hierarchical and k-means methodologies, customer clusters were refined.

Results and Interpretation: The analysis unveiled three discernible customer clusters: Loyal, Promising, and At-risk. Loyal patrons manifested recent, frequent, and substantial value transactions. Promising customers exhibited recent transactions but lacked frequency and high monetary value. At-risk customers demonstrated a dearth of recent transactions. Tailored recommendations were formulated for each cluster.

Implementation and Deployment: The research's insights materialized into strategic actions, concentrating efforts on Loyal and Promising customer clusters. This involved direct marketing initiatives, personalized campaigns, and optimized content delivery. Additionally, the company undertook eco-conscious packaging enhancements and introduced enticing combo offers.

The RFM analysis-driven customer segmentation proved instrumental in guiding Yeka Herbals Pvt Ltd towards more effective marketing strategies, customer engagement, and sustainable product offerings. The study's outcomes underscore the potency of data-driven decision-making in fostering business growth and enhancing customer relationships.

Keywords: RFM Analysis, Customer Segmentation, Marketing Strategies, Sustainable beauty products

PRODUCT MATCHING FOR ENHANCED COMPETITIVE ANALYSIS

Animesh Choubey

Sr. Data Scientist, Competitive Intelligence
Lowe's Services India Pvt Ltd
Bangalore, India
animesh.choubey@lowes.com

Nidhi Talwar

Data Scientist, Competitive Intelligence
Lowe's Services India Pvt Ltd
Bangalore, India
nidhi.s.talwar@lowes.com

Effective product catalogue matching across competitive companies has emerged as a critical imperative in today's dynamic business landscape, further fuelled by the ubiquitous "endless aisle" paradigm of online retail. This paper addresses the intensifying significance of this challenge by proposing practical solutions derived from real-world projects. The complexity of the task arises from diverse factors, including the multitude of products, disparate data sets, incomplete information, and varying data quality. Moreover, the perpetual introduction of new products complicates efforts to achieve precise categorization and alignment. With companies now grappling with the intricate task of maintaining pricing parity across millions of items to meet market demands and customer preferences, robust product matching solutions are indispensable.

In response, this paper concentrates on two pivotal questions central to successful product matching. Firstly, it delves into the categorization of product relationships, involving the intricate process of determining whether two products are identical, similar, or distinct. This nuanced process is quantified through a strategic amalgamation of natural language processing (NLP) techniques and neural network-based classification models. Secondly, the paper explores the strategic alignment of products from distinct company catalogues, a prerequisite for comprehensive comparative analysis. This alignment leverages retrieval techniques like BM25+ to efficiently identify initial product pairs for further investigation.

Key methodologies encompass supervised learning for precise product similarity quantification, deep learning techniques for cross-company catalogue matching, and the application of retrieval strategies for meticulous pair selection. Through these methodologies, the proposed approach enhances the accuracy of product matching, providing businesses with valuable insights to refine their product positioning and strategies, particularly in relation to aspects like price optimization. The practical implementation of these techniques further emphasizes their potential and effectiveness. Ultimately, this paper serves as a valuable guide for researchers, practitioners, and industry experts keen on enhancing catalogue matching across competitive enterprises, grounded in tangible real-world experiences.

Keywords: NLP Techniques, Retrieval Strategies, Neural Networks, Product Similarity, Real-World Implementation

HUMAN-IN-THE-LOOP FOR ACTIVE LEARNING AND RLHF

Kannan AS

Chief Digital Officer

NextWealth Entrepreneurs Pvt Ltd

Kalipatti, India

kannan.sundar@nextwealth.com

Deep Learning solutions are increasingly moving to production in the past few years with more enterprises going for widespread deployment. But the inherent challenge in going to production is the possibility of model drift once it goes to production. This can result in errors in the deployment leading to loss of credibility and customer attrition. This is inevitable because of the lack of availability of all the data upfront. Also, unlike traditional software development, maintaining the deployed model is hardly the job of a software engineer or a data scientist. This involves regular evaluation of the model performance manually and bringing in additional new data for continuous Active Learning of the model. This Human evaluation is critical to keep the model performing well in the field, especially if there are multiple deployments.

The new generation of Large Language Models have a need to be fine-tuned / instruction tuned for specific applications. Many of the implementations use RLHF (Reinforcement Learning with Human Feedback) to train large LLMs for specific applications, just as ChatGPT was implemented based on GPT3.5. Continuous human evaluation on the decisions taken by the finetuned model is a key driver for the performance of the Reinforcement Learning model.

On a different note, there is an increasing concern of loss of jobs due to deployment of AI in workplaces, which is not entirely unfounded, especially after dawn of the Large Language Models. This paper addresses a mechanism of handling the entire active learning lifecycle including live human-in-the-loop support for model errors with workforce from small towns. This paper also addresses some of the challenges of RLHF and how it can be addressed with Human-in-the-loop supported by Technology. The paper also highlights the social impact generated by this work being done in small towns across India.

Keywords: HIL, Active Learning, RLHF, Generative AI, Human Evaluation

CYBER CRIME: ANALYSIS AND PREDICTION USING MACHINE LEARNING ALGORITHMS

J Dharaneesh

Student

Coimbatore Institute of Technology
Coimbatore, Tamil Nadu
71762132012@cit.edu.in

S N Sasika

Student

Coimbatore Institute of Technology
Coimbatore, Tamil Nadu
71762132040@cit.edu.in

M Abinaya

Student

Coimbatore Institute of Technology
Coimbatore, Tamil Nadu
71762132001@cit.edu.in

S Deivarani

Assistant Professor

Department of Computing (Data Science)
Coimbatore Institute of Technology
Coimbatore, Tamil Nadu
deivarani@cit.edu.in

Dr. K Rajarajeshwari

Assistant Professor

Department of Computing (Data Science)
Coimbatore Institute of Technology
Coimbatore, Tamil Nadu
rajarajeshwari@cit.edu.in

Cyber Crime is a threat to the security of any individual, a community or to nation as a whole. Cyber Crime analysis and prediction is an approach which is used to find the patterns or trends that occur. Nowadays the numbers of cyber crime are increasing at an alarming rate. In the current scenario of rapidly increasing crime rate, a nation needs to analyse and predict the cyber crime before it occurs. To achieve this, few Machine Learning Algorithms are suggested. This paper studies about the cyber crime dataset of NCRB 2021 and provides the insights on crimes that occur state wise. The data set contains various attributes related to cyber crime tampering computer source documents, computer related offences, Ransomware, Dishonestly receiving stolen computer resource or communication on device, etc. Support Vector Machine and Random Forest algorithms are employed for classification and prediction. The main aim of this analysis is to help the authorities to detect, prevent and solve the cyber crime at an accurate and faster rate.

Keywords: Cyber Crime, Ransomware, Support Vector Machine, Random Forest

A DECISION TREE BASED PREDICTIVE MODEL TO CUSTOMER ACQUISITION IN ONLINE RETAIL

Dr. Vikas Kumar

Finance Officer
Central University of Haryana
Jant-Pali, Mahendergarh,
Haryana – 123031, India
prof.vikaskumar@gmail.com

Dr. Pooja Nanda

Associate Professor
Amity University,
Sector-125, Noida,
Uttar Pradesh 201301, India
poojan73@gmail.com

The Success of online retail business is dependent upon its capacity to attract, acquire, engage and convert consumers into customer, and customers into loyal customers. Therefore, it is vital that e-tailers know the numerous factors necessary to convert a consumer to a loyal customer with just a single visit to their portal. On the basis of literature review, this study has identified the important variables capable of driving customer acquisition and the necessary factors for providing an effective engagement/interaction with the system to facilitate conversion. It suggests how these variables can be incorporated into the website architecture and improved upon from time to time so that the business objectives can be achieved. Using the identified factors and variables, the study further proposes models for facilitating the acquisition of customers. In this model, we have utilized information collected from recent online visitors about gender, age, Location, shopping experience, shopping frequency, internet access, type of device, screen size, operating system, browser, channel, reaching mode, purchase time, payment option as an input and finally whether purchase made (get converted in to a customer) as a response variable. Data collected for 912 customers about their recent visit to online purchase has been used for the model. Using this information, we have tried to predict possibility of customer acquisition. Decision tree algorithm using R is applied to get insight and it is found that browser, reaching mode, purchase time, payment option and operating system are significant indicators to identify customer conversion. The proposed model is found to be very powerful as it has just one percent of misclassification error when trained model is tested on validation data. s In addition to above model we have also fitted random forest model to identify the variable importance as converting the online visitor in to a customer.

Keywords: Customer acquisition, user engagement, web Analytics, website quality, traffic source, online retail, demographic variables

A LIGHT-WEIGHT DEEP LEARNING MODEL FOR HAND GESTURE RECOGNITION IN SPACE APPLICATIONS

Sandeep Nithyanandan¹, Chithra A V², Somoshree Datta³, and Manusubramanian S⁴

¹ sandeepnithyanandan@lpssc.gov.in

² av chithra@lpssc.gov.in

³ somoshreedatta@lpssc.gov.in

⁴ manuss@lpssc.gov.in

Indian Space Research Organisation, Thiruvananthapuram, Kerala

Hand gesture recognition is a field of research involving computer vision and artificial intelligence to recognize and interpret hand gestures. Based on this interpretation, systems can be built and configured to perform various actions as needed. The research that has been carried out till date regarding the recognition of hand gestures involving space applications is very limited. Hence, we try to address this field of research whereby it is imperative to have accurate recognition of gestures for controlling space systems. The astronauts on board operate in very constrained environments. Both RGB and thermal data play vital roles in space exploration and enable astronauts to make informed decisions, troubleshoot problems, and enhance their understanding of the space environment. Hence, it is necessary to innovate and design precise AI models those will help in the smooth functioning of space systems. Moreover, in space missions, where resources like power, memory, and computational capacity are limited, it is crucial to have lightweight machine learning or deep learning models onboard. The architectural model that we propose in this paper is a light weight deep learning model which works on both RGB and Thermal modalities with better accuracy as compared to the state-of-the-art research. Our dual-mode model has achieved an accuracy of 93.27% on RGB images and 82.48% on thermal images respectively.

Keywords: Hand gesture recognition · lightweight · deep learning · deep CNN

UNLEASHING CUSTOMER INSIGHTS AND AMPLIFYING BUSINESS GROWTH: A CASE STUDY OF 'FAMILY IN MART' SUPERMARKET

Shanmuga Priya G

Post Graduate Diploma in Data Science and Business Analytics
Thiagarajar School of Management
Madurai, India
shanmugapriyagv23@gmail.com

Dr.Natraj Balasubramanian

Faculty Supervisor
Assistant Professor, Chairperson MDP
Thiagarajar School of Management
Madurai, India
bnataraj@tsm.ac.in

This study investigates 'Family in Mart,' a supermarket, aiming to comprehend its services, customer behavior, and operational efficiency. Employing RFM analysis and market basket analysis, it offers actionable insights for decision-making, enhancing customer experience, and optimizing services. RFM analysis employs SPSS software and clustering techniques, while the Apriori algorithm drives market basket analysis. The study uncovers distinct customer segments, facilitating targeted marketing, and reveals product associations for improved inventory management and cross-selling. Despite limitations related to single-supermarket focus and real-world data constraints, the findings have practical implications for tailored marketing, inventory management, and customer engagement, demonstrating innovative datadriven approaches in supermarket optimization.

Keywords: Customer segmentation, Marketing strategies, Market basket analysis, RFM analysis, Supermarket analysis

PREDICTING VOLUME MARKET SHARES OF CONSUMER BRANDS USING BAYESIAN DIRICHLET REGRESSION

Smit Rohan

Solutions Manager

Cogitaas AVA

Mumbai, India

smit.rohan@cogitaas.com

Volume market share of a consumer durable brand at any time t is defined as the ratio of the volume sales of the brand at t and the total volume of the category at t . This is a key metric to understand its relative standing amidst competition. The value of a market share is between 0 and 1 and the aggregated market shares of all brands within a category is 1 at any given time t . Hence, market share data is compositional in nature. Modelling and applying compositional data have been discussed extensively in the literature during the past 40 years and there are vector-based methods such as VARMA and dynamic time varying methods. A more recent entrant for modelling market shares is using Dirichlet regression. In our talk, we apply Bayesian Dirichlet regression for predicting market share of brands in a condiment category and discuss extensions to a dynamic setting.

Keywords: CPG, Dirichlet Regression, Market Share Model, Compositional Data, Time Series

REVIEW OF SENTIMENT ANALYSIS STUDIES USING SOCIAL MEDIA DATA

Dr. Kalpana S Kumaran

Associate Professor
ITM Business School
Navi Mumbai, India
kalpanas@itm.edu

Mr. Amit Sushil Dass

Student
ITM Business School
Navi Mumbai, India
2022.amitdass@itm.edu

This paper provides a summary and an examination of the analyses and findings from various studies on sentiment analysis using data from social media. People express their views, critics, feelings, opinions as well as provide feedback in the form of writings, photos, videos, and audio. These data are transformed into meaningful information using sentiment analysis and make better decisions based on the output. Studies published between 2018 and 2023 from eminent and reliable databases like Science Direct, Scopus, Emerald Insight, Web of Science, Springer, and IEEE Xplore were taken into consideration for the review of this research. The study's primary focus areas were global news, business, politics, and healthcare. After the extensive screening of 87 papers 33 papers were chosen for the review process. From each article the objectives, trends, values, concepts, techniques used, and conclusions were considered as important components to review the papers and to map the relevance between them from the papers. Most of the articles used lexicon-based approach then machine learning approach irrespective of the topics of the articles and the platforms used for the data collection were Twitter, Facebook, Reddit, Instagram, News forum blogs and community portal.

Keywords: Sentiment analysis, Social media, Lexicon-based, Machine learning

MARKET BASKET ANALYTICS DEFINE CUSTOMER BEHAVIOUR THROUGH TRANSACTION DATA

Dr. Kalpana S Kumaran

Associate Professor
ITM Business School
Navi Mumbai, India
kalpanas@itm.edu

Ms. Gauri Surve

Student
ITM Business School
Navi Mumbai, India
2022.gauris@itm.edu

The objective of this research is to design an application which can be used to analyse the existing transaction data. The feature of the application is market basket analysis. It is significant to understand the ever-changing customer behaviour and to identify product relationships that are frequently purchased together by customers to enable cross-selling & upselling opportunities, rise product placement, increase recommendations, better inventory management, and optimize pricing tactics which will contribute to profit maximization. Transaction data has frequently been utilized as a research object to find new knowledge. Utilizing transaction data is a better strategy than using the standard, time- and moneyconsuming customer surveys. Transaction data from Kaggle, a free data portal was used to find the best method for the implementation of market basket analysis. The trail result showed that association rule method using Apriori algorithm works very well. This study also demonstrates that while completing any market basket analysis, time is a crucial consideration that should be considered. This application can be used not only in retail, also in telecommunications, healthcare and anywhere transactional data can expose item associations.

Keywords: Market basket analytics, Association rule, Apriori algorithm, Transaction data

PARSER LLM-POWERED GENERATIVE AI FOR ENHANCED REAL ESTATE PROPERTY SEARCH

Sumit Ranjan

Associate Manager-Data Science

Majid Al Futtaim

Dubai, UAE

sumit.ranjan819@gmail.com

Lanwin Lobo

Director Data Science

MResult Services

Bangalore, India

lanwinlobo@gmail.com

Divya Chembachere

Lead Data Scientist

MResult Services

Bangalore, India

c.divya819@gmail.com

This abstract presents a novel approach to revolutionize the real estate property search process using Large Language Models (LLMs) and advanced AI techniques. As the real estate industry continues to evolve, embracing technological innovations can significantly enhance the property search experience. This paper introduces an innovative method that harnesses the power of LLMs to create a more intuitive, efficient, and personalized property discovery process.

Traditional property search methods often involve rigid keyword-based queries that may not capture the full spectrum of user preferences. LLMs, such as the renowned GPT-3.5, excel in processing natural language and understanding context. By allowing users to express their criteria conversationally, the property search process becomes more user-friendly and effective. For instance, users can input queries in a more natural way, like "Find homes with 3 bedrooms and a backyard close to a reputable school."

The core strength of this approach lies in the contextual understanding capabilities of LLMs. These models can grasp complex queries, identify user intent, and provide tailored property recommendations. By analyzing both structured data (such as property details) and unstructured data (like property descriptions and images), LLMs can deliver results that align closely with user preferences.

The proposed implementation involves a two-step process. First, the LLM analyzes the user's input to extract essential attributes and preferences. Next, it conducts a comprehensive search across property listings, taking into account a broader range of factors beyond the standard criteria. This may include evaluating amenities, considering local market trends, and even assessing future development potential.

However, the adoption of this innovative approach presents certain challenges. Ensuring that the LLMs understand domain-specific real estate terminology and regional nuances is crucial

for accurate results. Additionally, safeguarding user data privacy and providing transparent insights into the workings of the AI algorithm are essential to building user trust.

In conclusion, integrating Large Language Models into real estate property searches has the potential to redefine the way individuals find their dream homes. By enabling more natural language queries and harnessing contextual understanding, LLMs bridge the gap between user intent and search outcomes. This paper advocates for the incorporation of LLM-powered property search platforms, enhancing the accuracy, efficiency, and personalization of property recommendations. By embracing this technology, the real estate industry can empower users in their quest for the perfect home.

Keywords: AI Large Language Models, LLM, AI, real estate, property search, contextual understanding, personalized recommendations, natural language queries, user empowerment, technological innovation

REZSEARCH: AI-DRIVEN RESUME PARSER

Varenya Vinay

Student

St. Aloysius College

Mangalore, India

varennyav2000@gmail.com

Wincel Glany Pais

Student

St. Aloysius College

Mangalore, India

wincelpais@gmail.com

In the rapidly evolving landscape of recruitment, "rezSearch" emerges as a pioneering AI-driven resume parser, poised to revolutionize candidate evaluation processes. The core feature of rezSearch centres on its innovative AI technology, designed to extract pivotal information from diverse resumes efficiently. This novel approach expedites candidate screening, enabling recruiters to focus on core qualifications swiftly and effectively. rezSearch's integrated functionalities transcend traditional parsing methods. The system provides a comprehensive overview of each candidate's profile by categorizing critical resume components such as contact details, skills, education, work history, and certifications. This systematic organization, coupled with libraries like Numpy, Pandas, OpenAI, and more, ensures accurate data extraction and processing. Moreover, the incorporation of OpenAI's capabilities elevates rezSearch's language understanding, enabling it to decipher nuances within resumes of varying formats and complexities.

A standout aspect of rezSearch lies in its commitment to user-centric design. The software presents a user-friendly interface that simplifies the parsing experience. Recruiters can navigate through parsed data effortlessly, swiftly evaluating candidates and making informed decisions. In conclusion, rezSearch amalgamates cutting-edge AI capabilities, advanced libraries, and an intuitive interface, presenting a dynamic solution that redefines resume parsing and ultimately enhances the recruitment process.

Keywords: rezSearch, OpenAI, resume, AI

PROPHECY OF CAR VALUE USING MACHINE LEARNING TECHNIQUE

Dr. Ekta Agrawal

Associate Professor, Department of Computer Science,
Shri Vaishnav Institute of Management
Indore, Madhya Pradesh, India
ektaagrawal4jan@gmail.com

Dr. Jitendra Jain*

Associate Professor, Department of Computer Science,
Shri Vaishnav Institute of Management
Indore, Madhya Pradesh, India
jitendra1974@gmail.com

Dr. Kshama Paithankar*

Professor, Department of Computer Science,
Shri Vaishnav Institute of Management
Indore, Madhya Pradesh, India
kshama.paithankar@gmail.com

Dr. Ajay B. Gadicha*

Associate Professor, Department of Artificial Intelligence and Data Science
P. R. Pote Patil College of Engineering & Management,
Amravati, Maharashtra, India
ajjugadicha@gmail.com

In this swiftly-moving world, managing our professional as well as personal lives have become quite hectic and if we don't have our own personal vehicle for transportation, life is a lot more hectic. To be on the safe side, one should have a more reliable and easy mode for transportation and a personal vehicle is always the best option. Having a car is very important for people these days as it gives a certain social status and also gives a certain extent of personal control to individual owning it. In some areas with low population, having a car becomes essential as it provides the only option for covering long distances in case of an absence of public transport. Old aged people, who have difficulties in walking or cycling to places, have driving the sole option for moving without being dependent. And for those that don't have enough resources to purchase a brand-new car, buying an old vehicle becomes a necessity and that too at a reasonable price.

The car manufacturing has been increasing swiftly over the years during past decade, with about 92 million cars that were manufactured in 2019. This provides a big boost for the market of old and used cars which are now coming up as a progressively growing industry. The recent entries of various websites and web-portals have fulfilled the requirements of customers up to some extent as they now know the present trends and scenario to get the market value of any old vehicle present in the market. Machine Learning has a lot of applications in real world scenario but one of the most known applications is the use of Machine Learning in resolving the prediction problems. The proposed idea being discussed here is based upon one among such applications. Deploying various Machine Learning Algorithms to build a statistical model based upon given data and features set to estimate the prices of used cars.

Keywords: Car Prediction, Machine learning, Attributes (Data), System Architecture.

IMPACT OF GENERATIVE AI ON STUDENT AND TEACHER EXPERIENCES IN ACADEMIC JOURNEY

Dilip Kumar R

Cloud Solution Architect

Oracle Bangalore, India

ravichandrandilip@gmail.com

Rahul Gupta

Cloud Solution Architect

Oracle Kolkata, India

rahulgupta.rg@gmail.com

The rapid inflow of information in academic settings has necessitated novel solutions to augment the learning experience. This paper focuses on the application of advanced content summarisation algorithms in combination with intelligent quiz generation, catering to both student and teacher personas. The primary objective is to streamline the absorption of complex study materials and facilitate effective assessment to aid, expedite and enhance the overall education experience.

From a student perspective, content summarization algorithms employ contemporary natural language processing (NLP) techniques to summarise huge educational data, while retaining its essence. This not only allows students to grasp key concepts quickly and efficiently, but it also helps to optimize their study time, and bolster their understanding of the subject matter. Furthermore, the integration of AI-driven quiz generation enhances the learning process by offering interactive assessments that engage students with varying levels of difficulty. These quizzes are dynamically generated based on the summarized content, enabling students to gauge their comprehension and identify areas that require further attention.

From the perspective of teacher, the paper explores intricate algorithms that analyse educational materials and create concise summaries with retention of crucial information. The algorithms consider semantic associations, contextual importance, and domain-specific terminologies to ensure accurate interpretation. In addition to this, it helps the teachers to design quizzes aligned with curriculum goals from study materials.

This presentation investigates into the technical intricacies of the content summarization algorithms, with application of various techniques such as abstractive and extractive summarization, chunking techniques, and attention-based mechanisms. Furthermore, it explains the process of generating diverse quizzes, including question format variation, answer choice randomization.

In conclusion, the process of transcription of content into text format, using text information to create content summarization, quiz generation via complex algorithms holds the potential to aid different personas in their academic journey, overall revolutionize the academic landscape. By catering to both student and teacher needs, this approach empowers learners with efficient study aids and supports educators with dynamic evaluation tools. The paper underscores the significance of striking a balance between comprehensive content coverage and concise representation, offering insights into the intricate algorithms that drive these innovations.

Keywords: Generative AI, Summarization, Quiz generation, Natural Language processing, Academic application

IOT AND DEEP LEARNING BASED AGRI 4.0 PRACTICES FOR SUSTAINABLE FARMING

THIRAVIAM K*

Ph.D. Research Scholar

Department of Computer Applications, University college of Engineering (Bharathidasan
Institute of Technology – Campus), Anna University,
Tiruchirappalli, India
akthiraviam@gmail.com

KARTHIYAYINI R

Assistant Professor

Department of Computer Applications, University college of Engineering (Bharathidasan
Institute of Technology – Campus), Anna University,
Tiruchirappalli, India
karthi1vijay@gmail.com

The integration of the Internet of Things (IoT) and Deep Learning (DL) in the context of Agriculture Industry 4.0 (Agri 4.0) is a promising approach to addressing the pressing challenges facing existing farming practices. The implementation of this approach has the capability to transform agriculture into a sustainable enterprise capable of meeting the increasing demands of a rapidly changing world. Agri 4.0 applications, coupled with IoT and DL models, may provide a way to do it. In this article, current adaptation challenges and the state-of-the-art in Agri 4.0 are addressed. The areas of focus include timely intervention in environmental control using IoT to collect and analyse various environmental factors. This data can then be used to automate farm management and DL-based image processing using advanced neural networks for the detection of plant diseases, pests, and weeds, as well as for transparent agri-product grading. Besides, there are still some challenges in bringing such technologies to implementation in the small-scale agri-sector. However, this article demonstrates how the integration of these technologies in the Industry 4.0-based agriculture sector can transform it into a sustainable agro-industry.

Keywords: IoT, Deep Learning, Agri 4.0, Sustainable farming.

ENHANCING COMMERCIAL VIABILITY AND RUSTWORTHINESS: CONTENT VALIDITY FOR LLM-GENERATED OUTPUTS

Thirunithesh P R

II Year MBA

Thiagarajar School of Management

Madurai, India

thirunithesh07@tsm.ac.in

Dr. Nataraj Balasubramanian

Assistant Professor

Thiagarajar School of Management

Madurai, India

bnataraj@tsm.ac.in

Purpose – The purpose of this study is to add content validity to the LLM generated outputs in the area of personalized content generation to increase its commercial viability. It is indeed to provide content validity to the output of LLM's to improve their trustworthiness which will eventually help organizations to confidently rely on LLM API based applications for content generation tasks.

Design/Methodology/Approach – This approach constrains the knowledge of the LLM to a user created knowledge base. Whenever new data is added to the knowledge base, the LLM API is integrated using Langchain to identify the source or reference of data. Reference is stored along with the data in form of vectors in a vector database. During content generation, semantic search is used to fetch only the text chunks relevant to the topic from knowledge base and the knowledge of LLM will be constrained to text chunks fetched during semantic search to provide the output along with reference.

Findings – The existing literature on semantic search focuses on document retrieval. This paper illustrates an approach to provide source reference to add credibility to the LLM generated outputs by combining the semantic search along with user knowledge base. It addresses hallucination issues with the help of effective prompt engineering and constraining the knowledge of LLM.

Research Limitations/Implications – This method will effectively provide content validity which will make the LLMs into reliable content generation models, But the knowledge sources provided by the user should be from a credible source or should have a credible reference.

Practical Implications: These models can be tailored and used for commercial content generation purposes and will help in automation of the content writing jobs. This will offer more credibility to the LLM generated outputs.

Originality/Value This paper provides a model-based approach to provide source reference to the outputs generated LLM models in the content writing domain.

Keywords: Large Language Model, Vector Embeddings, Semantic Search, Content Validity, Generative AI, LangChain, Hallucination, Content writing

ESTIMATING MOBILE PHONE CATEGORICAL PRICES USING LOGISTIC REGRESSION

Dhruv Jore

Student

STME, SVKM's NMIMS Deemed to be University

Indore, India

dhruvjore@gmail.com

Amey Dubey

Student

STME,SVKM's NMIMS Deemed to be University

Indore,India

ameydubeydin@gmail.com

Shubhangi Jore

Associate Professor

STME, SVKM's NMIMS Deemed to be University

Indore, India

shubhangi.jore@nmims.edu

Vikas Khare

Associate Professor

STME, SVKM's NMIMS Deemed to be University

Indore, India

vikas.khare@nmims.edu

The price of mobile phones has a significant impact on their market success and sustainability. It is crucial to estimate the price of a mobile phone before its release to ensure optimal marketing and competitiveness with other products in the market. This prediction requires the collection of data on existing mobile phones and the application of different algorithms to reduce complexity and identify the most critical selection features for comparison within the data. The use of machine learning techniques like forward selection and backward selection helps to find the best price with maximum specifications. In this study, we collected data on mobile phones' specifications and features like camera, video, processor quality, and material quality. The dataset is complex to collect due to the thousands of mobile phone releases each year. Therefore, we selectively choose features to reduce the dataset's complexity and get an estimate of the price to release the product in the market. The Logistic Regression predicts the mobile phone's price in three categories (low-priced, moderately-priced, and highly-priced).

Keywords: exploratory analysis, price categorization, ml modelling, logistic regression

ANALYZING CONTENT CREATION AND CONSUMPTION ON YOUTUBE THROUGH VISUAL ANALYTICS

Prof. Nikita Ramrakhiani

Assistant Professor- Business Analytics
Vivekanand Education Society's Business School
Mumbai, India
nikita.ramrakhiani@ves.ac.in

Ms. Hiya Bhatia

Student- Business Analytics
Vivekanand Education Society's Business School
Mumbai, India
hiya.bhatia2224p@ves.ac.in

Ms. Preety Gind

Student- Business Analytics
Vivekanand Education Society's Business School
Mumbai, India
preety.gind2224p@ves.ac.in

Mr. Ajinkya Kedare

Student- Business Analytics
Vivekanand Education Society's Business School
Mumbai, India
ajinkya.kedare2224p@ves.ac.in

In today's digital landscape, social media platforms hold a pivotal role in disseminating content, nurturing interactions, and shaping user preferences. This comprehensive exploration delves into prominent YouTube channels, assessing subscribers, video views, and upload frequency across diverse categories and countries. As of January 2023, YouTube stands as a global content powerhouse with 2.5 billion monthly and 1.2 billion daily active users, engaging with a billion hours of content daily. With a rapid influx of 500 hours of video uploaded each minute, it remains an unparalleled hub for content creation. The platform's impact is global, notably in countries like India, the US, and Indonesia, with languages like English, Hindi, and Spanish reigning supreme. Categories spanning music, entertainment, gaming, education, and news captivate users, who dedicate an average of 2 hours and 24 minutes per session, viewing around 100 videos per month. The financial clout of YouTube is underscored by its \$28.8 billion revenue in 2022. Intertwining YouTube metrics with social media engagement indicators reveals the intricate cross-platform dynamics and user behavioral patterns. This layer adds dynamism to the narrative of digital content consumption. The integration of sociodemographic factors provides a broader comprehension of the intricate interplay between content engagement and societal attributes. Revealing captivating patterns and trends, this analysis unveils the evolving terrain of YouTube's content creation and audience engagement. Diverse data visualizations, including multi attribute plots to depict channel distribution, popularity, and interaction dynamics have been used in this research. This amalgamation extends beyond content creators and marketers, offering researchers valuable insights into digital platform dynamics within the broader context of societal indicators and evolving social media trends. Temporal trends in video views and subscribers are effectively conveyed through line charts, illuminating growth trajectories. This analysis delves into social media trends, acknowledging

their symbiotic rapport with YouTube. Metrics and engagement indicators converge, shedding light on cross-platform influences and user behaviors. Regression analyses untangle the nexus between sociodemographics and YouTube engagement. Crucially, this research aimed at finding how analytics reflects in the domains of social media and contributes towards refining strategies, optimizing engagement, and honing targeting efforts and acquiring ways to maximize the synergy that magnifies reach and cultivates cross-platform expansion.

Keywords: Social Media Analytics , Click-through rate, Data Visualisation using Tableau, Machine Learning, Predictive Modeling

REVOLUTIONIZING RISK ASSESSMENT: A NOVEL MACHINE LEARNING APPROACH FOR MOTOR INSURANCE LIABILITY PREDICTION

Yogesh Selvaraj

Trivandrum, India

Yogeshsp12@gmail.com

Assessing Insurance Liability accurately is of paramount importance for insurance companies to maintain financial stability and reduce the losses. Manual liability prediction in the insurance industry faces challenges due to subjective judgments, inconsistent data interpretation, and timeintensive processes. Human biases and lack of real-time insights hinder accurate risk assessment, leading to delayed claims processing, and potential inaccuracies in determining liability. Traditional methods for determining liability often involve complex rule-based systems that struggle to adapt to the dynamic nature of insurance data. In this context, machine learning techniques offer a promising avenue to enhance accuracy and efficiency in predicting insurance liability thereby reducing the losses. This Research proposes an innovative solution that integrates text mining techniques with a machine learning classification model to enhance liability prediction accuracy while providing transparent model explanations. Also the Research presents a comprehensive study and implementation of a machine learning classification model designed to predict the liability of an insurance company related to a claim. The model leverages a diverse imbalanced dataset containing historical claims information, policy details, claimant profiles, accident characteristics and other relevant variables. By employing advanced data preprocessing techniques, feature engineering, text-mining and state-of-the-art classification algorithms, the model learns complex patterns and relationships within the data. The solution contains multiple steps in building a ML based solution to help the business to accurately identify the Liability which involves

- 1) Data Compilation and Preprocessing
- 2) Text Mining and Feature Extraction
- 3) Handling Real world Imbalanced Data
- 4) Model Development
- 5) Training, Evaluation & Hyper-Parameter Optimization
- 6) Performance Metrics and Explainability Evaluation
- 7) Model Explainability

Keywords: Machine Learning, Text Mining, Imbalanced Data, LightGBM

HEALTH ANALYTICS ON IBM Z LEVERAGING IBM SNAPML

Divya Prakash Kulkarni

Data Scientist

IBM

Bangalore, India

Divya.kulkarni1@ibm.com

Chandra Shekhar Reddy Potula

Lead AI on zSystems Solution Team

IBM

Bangalore, India

chandra.shekhar@in.ibm.com

In medical insurance organizations, the medical claims amount that is expected as the expense in a year, plays an important factor in deciding the overall financial growth the company. According to the Mckinsey Global Insurance Report 2022, Twelve insurers that write homeowners coverage in Louisiana were declared insolvent between July 2021 and February 2023. Insurers have paid out more than \$23 billion in insured losses from over 800,000 claims filed from the two years 2021 & 2022. Moreover, Health insurance stands as one of the most significant annual investments made by individuals, providing financial protection against the costs incurred due to disease, accidents, and illnesses.

In the health insurance industry, Artificial Intelligence (AI) has emerged as a transformative force, reshaping various aspects of healthcare operations and decision-making. The integration of AI brings about a paradigm shift in how insurers evaluate risks, optimize processes, enhance customer experience, and ultimately improve healthcare outcomes.

This Paper/Demo explains about implementation of Health Insurance use cases like Health Insurance Premium prediction and Claims prediction. The objective is to utilize IBM SnapML for the creation of predictive models in the realm of Health Insurance and subsequently deploy these models on IBM Z i.e. train anywhere your model but can deploy on IBM Z which can leverage the on chip AI Accelerator capabilities on IBM Z16. IBM SnapML offers very powerful, multi-threaded CPU solvers, as well as efficient GPU solvers and IBM Z16 is designed for analysing real-time transactions at scale and is industry-first quantum-safe system.

Keywords: AI on Mainframes, SnapML, IBM Z16, Health Analytics

PREDICTING CUSTOMER INTEREST IN VEHICLE INSURANCE

Dhruv Jore

Student

STME, SVKM's NMIMS Deemed to be University

Indore, India

dhruvjore@gmail.com

Amey Dubey

Student

STME,SVKM's NMIMS Deemed to be University

Indore,India

ameydubeydin@gmail.com

Shreyas Arora

Student

STME , SVKM's NMIMS Deemed to be University

Indore,India

shreyasarora0123@gmail.com

Vikas Khare

Associate Professor

STME, SVKM's NMIMS Deemed to be University

Indore, India

vikas.khare@nmims.edu

This research paper aims to predict customer interest in purchasing vehicle insurance from a health insurance company using data science techniques. The study uses data collected from health insurance holders of the company and employs data processing, analysis, and visualization techniques to prepare the data for machine learning models. Three models, namely logistic regression, Naive Bayes, and random forests, were trained on the data and their performance matrices were compared to determine the best-performing model. The study concludes that the random forests model outperformed the other two models in predicting customer interest in purchasing vehicle insurance. The results of this study have practical implications for health insurance companies seeking to expand their service offerings to customers. The study contributes to the growing field of data science in the insurance industry and highlights the potential benefits of using machine learning models to predict customer interests. The results of this research can aid insurance companies in developing targeted marketing strategies and improving customer satisfaction.

Keywords: Naive Bayes, random forest, logistic regression , machine learning

FORECASTING OF INDIAN WEATHER: VISIBILITY PREDICTION USING MACHINE LEARNING TECHNIQUES

Prof. Nikita Ramrakhiani

Assistant Professor– Business Analytics
Vivekanand Education Society's Business School
Mumbai, India
nikita.ramrakhiani@ves.ac.in

Mr. Harsh Patel

Student– Business Analytics
Vivekanand Education Society's Business School
Mumbai, India
harsh.patel2224p@ves.ac.in

Mr. Amey Hulsure

Student– Business Analytics
Vivekanand Education Society's Business School
Mumbai, India
amey.hulsure2224p@ves.ac.in

Mr. AbhijeetKumar Srivastava

Student– Business Analytics
Vivekanand Education Society's Business School
Mumbai, India
abhijeetkumar.srivastava2224p@ves.ac.in

Weather reports are essential for a range of purposes, including pollution control and addressing weather-related issues. They help identify the dispersion of pollutants, enabling effective pollution management. Weather forecasts also aid in preparing for emergencies, optimizing agriculture practices, ensuring safe transportation, and promoting public health. These reports play a vital role in decision-making across sectors, enhancing safety, resilience, and overall well-being. Using machine learning, particularly the logistic regression algorithm, this study demonstrates its efficacy in predicting binary weather events like rain or sunny based on various weather parameters. Through accurate modelling and evaluation, the research highlights the algorithm's potential for practical weather forecasting applications. This research intends to analyze the significance of weather forecasting for advancing our comprehension of Indian weather dynamics and its role in enhancing predictive capabilities to address weatherrelated challenges effectively. Through our model and analysis, we attempt to predict the visibility index and identify the factors such as pollution, gas emissions, etc which affect visibility in specific regions which impact the number of accidents and mishaps.

Keywords: Weather Forecasting, Air Pollution, Visibility, Machine Learning, Tableau

MACHINE LEARNING FOR AUTOMATED FLOOR DETECTION IN BUILDING IMAGES FOR THE INSURANCE SECTOR

Narendra Kumar

Project Manager

Allianz Services

Trivandrum, India

narendra.kumar1@allianz.com

Varsha Shaji

Associate Analyst

Allianz Services

Trivandrum, India

varsha.shaji1@allianz.com

Sangeeth G

Associate Analyst

Allianz Services

Trivandrum, India

sangeeth.g@allianz.com

Rahul Mahajan

Deputy Manager

Allianz Services

Trivandrum, India

rahul.mahajan@allianz.ie

In the insurance sector, precise assessment of building properties is crucial for determining appropriate coverage and insurance premium amounts. Effective risk evaluation, formulating policies, and determining premiums depend on the accurate assessment of the relevant properties of the building to be insured. One such critical property is the number of floors in a building. Traditional methods of floor count estimation, such as surveys and sending an investigator, are often time-consuming and quite manual. Further, such methods are prone to inconsistencies and errors, which can be detrimental to the overall decision process. This paper presents a solution for the above challenge by utilising machine learning (ML) techniques for automating floor count detection using building images. In this paper, we propose an object detection model that can accurately estimate the number of floors in a building from front-view images of the building. The proposed solution is a multi-step methodology comprising data preprocessing, feature extraction, and ML model training. It involves the creation of a diverse dataset of annotated building images curated by encompassing various building heights and architectural styles curated explicitly for this task, preprocessing the images, and training and comparing multiple ML models to determine the floor count. The effectiveness of the proposed model is demonstrated in the experimental results, which show that the model can accurately determine the number of floors across diverse building types. Implementing this automated floor count detection system can help simplify property assessment within the insurance sector, reducing cost, processing time, and manual effort. Different object detection algorithms are compared to determine the efficient model for detecting the floors. In future, additional features Internal will be considered in the property assessment. The solution mitigates human error and leads to more precise risk assessments.

Keywords: Building floor Detection, Object Detection, Deep Learning, Assessment

EXPLORING THE INTERSECTION OF MACHINE LEARNING AND VOICE PATHOLOGY: A COMPREHENSIVE REVIEW

Divya O M

Research Scholar

Department of Computer Science

CHRIST University

Bangalore, India

divya.om@res.christuniversity.in

Dr Sagaya Aurelia

Assistant Professor

Department of Computer Science

CHRIST University

Bangalore, India

sagaya.aurelia@christuniversity.in

Voice disorders may have a significant impact on an individual's quality of life and are often associated with occupational, social or health challenges. In recent years, the integration of machine learning techniques helps in the field of voice pathology to make significant advancements. This comprehensive review aims to explore the multifaceted relationship between machine learning and voice pathology, shedding light on how these innovative technologies are revolutionizing the diagnosis, assessment, and understanding of voice disorders. One of the pivotal applications of machine learning in voice pathology is automated diagnosis. The paper outlines various machine learning models which have been deployed to distinguish between healthy and pathological voices.

In this paper, a study is made on different machine learning techniques as well as feature extraction methods used in the literature for defining the characteristics of the voice and how they are used as input for machine learning algorithms/analysis for predicting the pathological voice from the normal one. A thorough examination of diverse machine learning techniques and feature extraction methods utilized in the realm of voice analysis will give insight into how these methods contribute to characterizing vocal traits, ultimately serving as inputs for machine learning algorithms. The central focus lies in using these approaches to effectively discern between pathological and normal vocal conditions. Through a meticulous study of the literature, we aim to provide valuable insights into the advancements and challenges within this domain.

The review addresses limitations such as the need for standardized datasets, interpretability of machine learning models, and the ethical implications of AI-assisted diagnosis. It also highlights ongoing research directions, including the integration of real-time monitoring devices to enhance the trustworthiness of machine learning-based diagnostic outcomes.

Based on this study, it could be observed that classification and prediction techniques could be efficiently applied over the voice for pathology prediction with more accuracy and less human intervention.

Keywords: Machine Learning, Voice Pathology, Feature Extraction, Classification, Prediction

CAR PRICE INFLUENCE ANALYSIS: FEATURES

Ganesh Balaji R

Department of Computing
(Student – MSc Data Science)
Coimbatore Institute of Technology
(Affiliated to Anna University)
Coimbatore, India
71762132015@cit.edu.in

Ms Deivarani S

Department of Computing
(Assistant Professor)
Coimbatore Institute of Technology
(Affiliated to Anna University)
Coimbatore, India
deivarani@cit.edu.in

Kavin Chant P

Department of Computing
(Student – MSc Data Science)
Coimbatore Institute of Technology
(Affiliated to Anna University)
Coimbatore, India
71762132018@cit.edu.in

Dr Rajarajeshwari K

Department of Computing
(Assistant Professor)
Coimbatore Institute of Technology
(Affiliated to Anna University)
Coimbatore, India
rajarajeshwari@cit.edu.in

In the ever-evolving Indian automobile market, understanding the driving factors behind car prices is crucial for manufacturers to refine product offerings and tailor marketing strategies effectively. This machine learning project aims to uncover the most influential features impacting car prices, assisting manufacturers in making informed decisions. The project leverages a comprehensive dataset containing over 1200 car models/variants, encompassing a wide range of features such as engine specifications, fuel efficiency, brand, model, and variant details. A comprehensive analysis was embarked upon to unveil the features that significantly influence car prices. The dataset was pre-processed, handling missing values and converting categorical variables into numerical representations for further analysis. The core of the project involves the development of predictive models for car prices, employing both Decision Trees and Random Forests regression algorithms. These models excel in capturing complex relationships between features and prices. Following the training process, the feature importance scores were extracted from each model. Visualizations, such as bar plots and heatmaps, aid in interpreting the results, highlighting the features with the highest importance scores. By systematically analyzing these insights, manufacturers gain a comprehensive understanding of the features that wield the greatest influence on car prices. Armed with this

knowledge, they can prioritize improvements to specific attributes, adjust marketing campaigns to spotlight influential features, and better cater to evolving customer preferences. In conclusion, unveiling the feature importance enable the manufacturers to enhance their competitive edge by offering products that resonate with customers and maximizing their market presence. Through a combination of advanced machine learning techniques and insightful visualizations, this project contributes to a more informed and adaptive automotive sector.

Keywords: Indian automobile market, Machine learning, Feature importance, Predictive models, Marketing strategies

INTEGRATING REGRESSION AND TIME SERIES FORECASTING FOR ENHANCED TRAFFIC PREDICTION

Likkitha S

Student

Coimbatore Institute of Technology

Coimbatore, Tamilnadu

71762132004@cit.edu.in

Ajithram AS

Student

Coimbatore Institute of Technology

Coimbatore, Tamilnadu

71762132023@cit.edu.in

Mathan Kumar P

Student

Coimbatore Institute of Technology

Coimbatore, Tamilnadu

71762132024@cit.edu.in

S.Deivarani

Assistant Professor Department of Computing (Data Science)

Coimbatore Institute of Technology

Coimbatore, Tamilnadu

Dr. K.Rajarajeshwari

Assistant Professor Department of Computing (Data Science)

Coimbatore Institute of Technology

Coimbatore, Tamilnadu

Nowadays, congestion in traffic is increasing due to numerous vehicles flow on road during peak hours. So, predicting traffic system is essential for effective transportation management, congestion reduction, and infrastructure planning. This paper focuses on the techniques that would predict the traffic in order to avoid congestion and moreover waste of time in travelling. This paper presents a hybrid approach that combines regression analysis and time series forecasting to improve the precision of traffic prediction models. By integrating historical traffic data, temporal patterns, weather conditions, and relevant events, we construct a comprehensive dataset for training and testing. The regression component of the hybrid model presents the influence of exogenous variables such as weather and events on traffic patterns, while the time series forecasting component gives the inherent temporal dependencies within the data. In the regression phase, we employ techniques like linear regression and support vector regression to quantify the impact of external factors on traffic flow. Simultaneously, the time series forecasting phase employs methods like Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing to capture the temporal trends and cyclic patterns present in traffic data. To evaluate the performance of our hybrid approach, we conduct experiments using real-world traffic datasets from urban areas. Predictive accuracy is measured using metrics including Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and R-squared (R²). Comparison of the hybrid model against standalone regression and time series models is carried out to demonstrate its superiority in handling both exogenous

and endogenous factors. The finding shows that the hybrid model outperforms the individual techniques, offering a more comprehensive understanding of the traffic dynamics. The combination of regression and time series forecasting provides a prediction framework that considers both external influences and the historical patterns of the dataset. In conclusion, this paper highlights the efficacy of combining regression and time series forecasting for efficient traffic prediction. The hybrid model's ability to incorporate both exogenous and endogenous factors results in more accurate predictions. This helps the urban and traffic management authorities in making informed decisions to alleviate congestion and enhance the overall urban mobility.

Keywords: traffic prediction, regression analysis, time series forecasting, hybrid approach, urban transportation, congestion reduction, infrastructure planning

INK TO INSIGHT: SVMS AND NEURAL NETWORKS FOR PRECISE DIGIT RECOGNITION

Shreemathi M

Student

Department of Computing (MSc Data Science)

Coimbatore Institute of Technology

Coimbatore, India

71762132044@cit.edu.in

Sunitha S

Student

Department of Computing (MSc Data Science)

Coimbatore Institute of Technology

Coimbatore, India

71762132051@cit.edu.in

Yashvitha P R

Student

Department of Computing (MSc Data Science)

Coimbatore Institute of Technology

Coimbatore, India

71762132058@cit.edu.in

In the realm of modern pattern recognition, accurate handwritten digit detection holds paramount significance for applications like character recognition, document processing, and data entry automation. This project introduces an innovative approach that synergizes Support Vector Machines (SVMs) and neural networks within a unified framework, effectively enhancing digit detection capabilities.

Driven by the aim of precise digit identification within intricate datasets, this project seamlessly integrates the strengths of SVM models and neural networks to elevate detection accuracy while maintaining efficiency. SVMs are adept at classification tasks due to their ability to discern intricate data distributions, enabling the delineation of optimal hyperplanes for digit recognition. On the other hand, neural networks emulate human brain structures, enabling the capture of intricate data patterns and relationships essential for accurate digit recognition.

The project's workflow includes capturing digit images, preprocessing for data refinement, and careful dataset partitioning. SVM models undergo rigorous training to identify digit features and relationships, while neural networks are tailored for nuanced digit recognition. The culmination involves a comprehensive evaluation of both models using metrics like accuracy and precision.

Therefore, this project offers a promising avenue for advancing digit detection through the integration of SVMs and neural networks. This synthesis holds potential to refine digit recognition, fostering accuracy and utility across various applications.

Keywords: Handwritten digit detection, Support Vector Machines (SVMs), neural networks, accuracy enhancement

RECURRENT NEURAL NETWORK BASED TEMPERATURE DRIFT COMPENSATION IN TFG

Anjitha Viswanath

Scientist/Engineer
ISRO Inertial Systems Unit
Thiruvananthapuram, India
anjithavssc@gmail.com

Tony Thomas

Scientist/Engineer
ISRO Inertial Systems Unit
Thiruvananthapuram, India
tony_thomas@vssc.gov.in

T Sivasankar

Division Head
ISRO Inertial Systems Unit
Thiruvananthapuram, India
t_sivasankar@vssc.gov.in

Owing to the recent development in machine learning algorithms and technology, neural networks have a wide range of applications in diverse domains. The prediction and compensation of temperature induced drift in gyroscopes using neural networks is highly explored because of its adaptive configuration and the ability to handle non-linear problems.

Tuning fork gyroscope (TFG) is a Coriolis vibrating gyroscope (CVG) which works on the Coriolis principle and uses forks as vibrating elements to measure the angular rate. These are installed in inertial measurement units along with the accelerometers to determine the navigation states in land, aerospace and space applications. They are reliable, simple and cheaper with lower mass and power consumption. However, the drift in gyro bias with ambient temperature can lead to degradation in navigation accuracy.

Classical piecewise and polynomial regression models being unable to predict and compensate this drift in bias, the paper addresses the use of neural networks (NN) for the same. Although, literature states that memory less neural networks help in bias compensation under static condition, it is not effective during thermal gradients. Hence, the study focuses on the implementation of memory based neural network models namely, simple recurrent neural network (RNN), long short term memory (LSTM) NN and gated recurrent unit (GRU) NN.

RNN is unique with its built-in memory capability while handling time series data. However, since it uses gradients to update the weights, if the gradient values are too small, it deteriorates the learning and hence it forgets the past inputs in longer sequences resulting in vanishing gradient or short term memory problem. LSTM and GRU address this issue by using its internal gated mechanism. The gating mechanism enables the network to identify and keep only the relevant data in the long sequence data chain. Further, GRU with reduced number of gates offers faster computation and lesser implementation complexity in comparison to LSTM.

For the generation of temperature compensation model using recurrent, LSTM and GRU neural networks, a methodology was identified and finalized for acquiring the data sets for training,

validation and test purposes. The data sets were generated for various sensor parameters under thermal cycling condition with different thermal gradients and also under static room conditions. The input features (frequency, amplitude and quadrature) were also decided upon. Further, the training parameters (loss function, optimiser, batch size and patience) and the model parameters (size, layers, memory and activation function) were optimized. Results proved that the temperature compensation models generated using memory based neural network prediction provided more than 70% improvement in the sensor performance.

Keywords: Recurrent neural network (RNN), Long short term memory (LSTM), Gated recurrent unit (GRU), Tuning fork gyroscope (TFG), Temperature drift compensation

ANALYSIS OF ROI AND ESG IN OIL AND GAS SECTOR

Priyanshi Jain

Student

Welingkar Institute of Management,
Development and Research

Mumbai, India

1rba2022-priyanshi.jain@welingkar.org

Rishabh Kumar Arora

Student

Welingkar Institute of Management,
Development and Research

Mumbai, India

1rba2022-rishabh.arora@welingkar.org

Dr. Tanushree Bhattacharjee

Associate Professor

Welingkar Institute of Management,
Development and Research

Mumbai, India

tanushree.bhattacharjee@welingkar.org

Dr. Kavita Kalyandurgmath

Associate Dean

Operations and program In-charge-Research & Business Analytics

Welingkar Institute of Management,
Development and Research

Mumbai, India

kavita.kalyandurgmath@welingkar.org

ESG, encompassing Environmental, Social and governance factors, has emerged as a pivotal framework for organizations aiming to balance Sustainability and Ethical impact. Recognized by prominent firms like McKinsey, Deloitte, and KPMG, ESG's significance lies in its ability to drive financial performance, mitigate risks, and enhance reputation. This framework resonates globally, with industries like Oil and Gas, as seen in India's Directorate General of Hyderabad's report, acknowledging ESG's role in addressing climate change, environmental protection, social responsibility, and effective governance.

This research narrows its focus to the Environmental dimension of ESG, delving into vital factors like Sulphur Oxide Emissions, Total Water Withdrawal, and Supply Chain Environmental impact etc. By establishing a correlation between these Environmental metrics and financial parameters, the study unveils the intricate relationship between Sustainability efforts and Economic Outcomes. A comparative analysis spanning the Pre-Covid and Post-Covid periods accentuates the profound impact of pandemic-induced restrictions on environmental scores. The unprecedented lockdown measures, prompted a stark divergence in results, underscoring how external disruptions can reshape companies' ecological footprint and ESG performance.

In this comprehensive research, a robust methodology is employed, combining extensive data sources to analyse, the Environmental facet of ESG within the oil and gas sector. A multitude of Environmental Reports sourced from Bloomberg, in conjunction with data extracted from Annual Reports, form the bedrock of Quantitative analysis. Furthermore, a Qualitative dimension is introduced through Primary research, involving insightful interviews with key officials from prominent players in the industry such as HPCL, ONGC, BPCL, among others. This triangulation of data sources not only enhances the accuracy of findings but also offers a holistic perspective on the intricate relationship between environmental performance and financial parameters.

Employing a sophisticated technological toolkit encompassing Excel, Python, and Power BI, this research transcends conventional analysis boundaries. These tools facilitate meticulous data processing, visualization and pattern recognition, enabling a comprehensive evaluation of environmental parameters and their financial implications in the oil and gas domain. Beyond its analytical prowess, the research aspires to catalyse a transformative shift in investment paradigms. By shedding light on the intricate interplay between environmental factors and financial outcomes, it empowers stakeholders to make informed, holistic and sustainable investment choices.

Keywords: ESG Score, Correlation, Pattern Recognition, Data Visualisation, Oil and Gas Sector, Financial Parameters

PREDICLEND : KNOW YOUR LOAN BEFORE YOU APPLY, WHERE ACCURACY MEETS ACCESSIBILITY

Shaun Jose C

Department of Computing
(Student- Data Science)
Coimbatore Institute of Technology
Coimbatore, India
71762132043@cit.edu.in

Ms.Deivarani.S

Department of Computing
(Assistant Professor Data Science)
Coimbatore Institute of Technology
Coimbatore, India
deivarani@cit.edu.in

Sriram P

Department of Computing
(Student- Data Science)
Coimbatore Institute of Technology
Coimbatore, India
71762132048@cit.edu.in

Dr. Rajarajeshwari.K

Department of Computing
(Assistant Professor-Data Science)
Coimbatore Institute of Technology
Coimbatore, India
rajarajeshwari@cit.edu.in

In the modern era of digital banking, the ability to efficiently and accurately assess loan eligibility is paramount for financial institutions. This project presents a Machine Learning(ML) solution designed to predict whether an individual is eligible for a loan based on a set of predictor variables, including age, salary, and marital status. This project provides a user-friendly front-end interface integrated with a bank's website, enabling users to input their details and receive an instant assessment of their loan eligibility. Users are prompted to provide their information, including age, salary, and marital status, through a user-friendly web form. Upon submission, the integrated ML models process the input and swiftly generate predictions of the user's loan eligibility. These predictions are communicated back to the user via an intuitive interface, enhancing the user experience and fostering greater transparency in the loan application process. This project showcases the successful implementation of a predictive loan eligibility assessment system.

Keywords: Interactive User Experience, Data-Drive

EMAIL CAMPAIGN ANALYTICS FOR GENERATING LEADS: A CASE STUDY OF A CONSULTING FIRM

Sushant Kumar Mishra

Research Student

Research & Business Analytics

Prin L.N Welingkar Institute of Management Research and Development, Mumbai
Mumbai, India

sushant1991km@gmail.com

Dr. Aradhana Goutam

Associate Professor

Research & Business Analytics

Prin L.N Welingkar Institute of Management Research and Development, Mumbai
Mumbai, India

aradhana.goutam@welingkar.org

Shashank Prabhat

Research Student

Research & Business Analytics

Prin L.N Welingkar Institute of Management Research and Development, Mumbai
Mumbai, India

shashankprabhat442@gmail.com

Dr. Kavita Kalyandurgmath

Associate Dean-PGDM(Ops), Incharge

Dean- Research & Business Analytics

Prin L.N Welingkar Institute of Management Research and Development, Mumbai
Mumbai, India

kavita.kalyandurgmath@welingkar.org

This research explores a wide array of implementation factors within a subset of opt-in email marketing campaigns in the consulting firm. The dataset included approximately 2249 promotional email distributed over a duration of 2 months. An analysis of the email campaigns content uncovers diverse strategies encompassing layout, recipient addressing, subject lines, and interactive elements. These strategies are employed with the dual purpose of initially captivating recipients' interest and subsequently nurturing deeper engagement. This study tries to pinpoint the key strategies for attracting and keeping people through email campaigns. So, during email campaign, the following key metrics are analyzed: Open Rate (9%), Conversion Rate (10%), Bounce Rate (40%), Time and Day Analysis (Best day-Thursday; Best time: (12pm to 1 pm), A/B Testing Results, List growth rate (20%), Geographic Insights (Mumbai), etc. This study's findings specify the insights which are required to continuously refine the email campaigns and precise steps that must be taken to draw in and keep customers that rely on the Internet's digital marketing. Also shown are the time and day analysis in this study which illustrates that by tracking when emails are opened and interacted with, a future campaign for maximum interaction can be planned.

Keywords: Email-driven sales funnel, Lead Generation Analytics, B2B Sales lead generation, Conversion-focused email marketing, Email - based prospecting

GENERATIVE AI BASED SOLUTION FOR SALES AGENT'S PRODUCTIVITY IMPROVEMENT IN LIFE INSURANCE INDUSTRY

Sanjay Thawakar

Senior Vice President

AI works & BPMA Max Life Insurance

Vibhu Srivastava

Vice President

AI works Max Life Insurance

Divyan Kavdia

Vice President

AI works Max Life Insurance

Abhishek Kumar

Senior Specialist

AI works Max Life Insurance

Abhilash Pujar

Senior Manager

AI works Max Life Insurance

Keywords: Generative AI, Large Language Models, Lang chain, Embedding models, vector database and text retrieval, Insurance sales productivity management, Insurance sales training

The Life insurance policy sales in India is primarily driven by the insurance sales agents, either through financial sales agencies or through banks which have partnered with Life Insurance companies to offer LI products to their customers. While the Life Insurance penetration in India of ~3%, and a rapid industry growth rate of ~30%, presents a huge opportunity for the industry, the Life insurance sales operations in India present significant challenges driven by multiple factors such as the complexity of the Life Insurance products itself, the regulations and monitoring around sales process and low financial literacy and awareness about the Life insurance products among potential customers.

Apart from managing the day to day sales activities, Life insurance companies deploy large number of highly skilled sales training and capability development teams which enable continuous recruiting and training of newly hired sales agents (10,000-50,000 for large LI companies) along with training the entire sales teams on every new life insurance product launched in the market. In recent times, the role of these sales training teams have become highly significant and critical due to high magnitude of agent level attrition (ranging from 30-50% driven by industry competition) and higher frequency of new complex product launches in the LI industry.

With limited availability of highly skilled sales training teams, the Life insurance companies face significant challenges wrt incurring high cost of sales operations due to lower availability of trainers, which makes it difficult to provide timely and focused guidance and support to new sales team members beyond the initial classroom trainings. This directly impacts the company sales with lower proportion of agents becoming sales productive within even after 6-12 months of joining the sales force.

Research Objective

The purpose of the research was to build an end to end generative AI (Large Language model) based conversational chat application, which can act as an intelligent virtual trainer and sales assistant to the on-field sales agents reducing the dependency of on ground sales agents on skilled trainers and experienced sales professionals. This chat application acts as a question answering bot interacting in natural, human-like conversations with the right context, tone, and intent of the query in multiple languages, to provide answers to all the sales and product related questions to every sales agent on the mobile anytime, anywhere.

Approach

The overall approach involved leveraging various open source Large Language models such as– GPT 3.5 turbo along with techniques such as supervised LLM fine tuning and prompt tuning using the Max Life specific agent sales training and product specifications documents and training materials (text, video, speech-audio) to develop a question answering chat bot to provide on-demand, real-time and interactive responses for the sales agents.

Architecture

The solution architecture involved curating and creating transcriptions of the sales training materials (500+ documents) from various classroom training presentations (PowerPoint, PDFs), video recording of sessions and audio podcasts using document loaders, leveraging text embedding models to create word embedding's of documents and store it in the form of vector database and retrieval of this information as question and answers, through a Lang chain orchestrator along with a chat interface for user interaction.

Outcomes

The outcome of the research was a state of the art solution-generative AI based chat-bot with capability to input any form of data – text, audio or video and generating the appropriate answers in Max Life specific context (Max Life products and sales training practices), required sales tone and persona to all sales related questions from the sales agents with comprehensive capability to provide –

- Detailed sales scripts basis the customer demographics, desired language and tone
- Analyze customer data to recommend accurate and suitable products to the customer basis stage of life and demographics
- Selecting the product variant and technical knowhow of each variant
- Sharing smart communication conversation starters for customer interactions
- Handling customer sales objections with generated real life examples and with supporting external links from the internet
- Competitor product comparisons and differentiators Current Affairs that affect the world of finance and influence purchase decisions

SPATIAL MODELING OF LANDSLIDE SUSCEPTIBILITY USING RANDOM FOREST TECHNIQUE

Lakshay; Additya Akash Mishra; Pratham Sharma

B.Tech. CSE Student

NIIT University

Neemrana, India

lakshay.chhikara20@st.niituniversity.in;

additya.mishra20@st.niituniversity.in;

pratham.sharma20@st.niituniversity.in

Thota Sivasankar

Assistant Professor

NIIT University

Neemrana, India

siva.iirs@gmail.com

Debashis Sengupta

Dean (Technology Initiatives) and Area Director (CSE & GIS)

NIIT University

Neemrana, India

debashis.sengupta@niituniversity.in

This study aims to develop a comprehensive landslide susceptibility map to assess the vulnerability of regions to landslide occurrences. The map relies on the integration of various influential natural factors termed as Landslide Causing Factors (LCFs), encompassing variables such as Rainfall, Slope, and other important aspects. Focusing on the Hindu Kush Region, we meticulously gathered data pertaining to these LCFs. While some data, including Rainfall, NDVI, LULC, and DEM, were directly sourced from Google Earth Engine, other factors like Slope Aspect, Convergence Index, Drainage Density, and Relative Slope Position were derived from primary factors.

To optimize the quality of the susceptibility map, a judicious factor selection process was executed. Recognizing that not all factors hold equal weight in landslide occurrences, we employed the Chi-square method to eliminate non-contributory variables. This step refined the pool of factors for subsequent analysis. To further enhance accuracy, a Multicollinearity Assessment was conducted, shedding light on the relative significance of each selected factor.

The challenge of this research lies in the utilization of the Random Forest machine learning technique, which facilitates the translation of anticipated outcomes into tangible representations. This technique acts as an automated and efficient framework for realizing landslide susceptibility mapping objectives. The resultant map serves as an important input for informing regional planning initiatives, effectively identifying areas of heightened risk and enabling the implementation of strategic projects.

Keywords: Spatial modelling, Landslide susceptibility, Machine learning, Hindu Kush Himalayas



Thank You