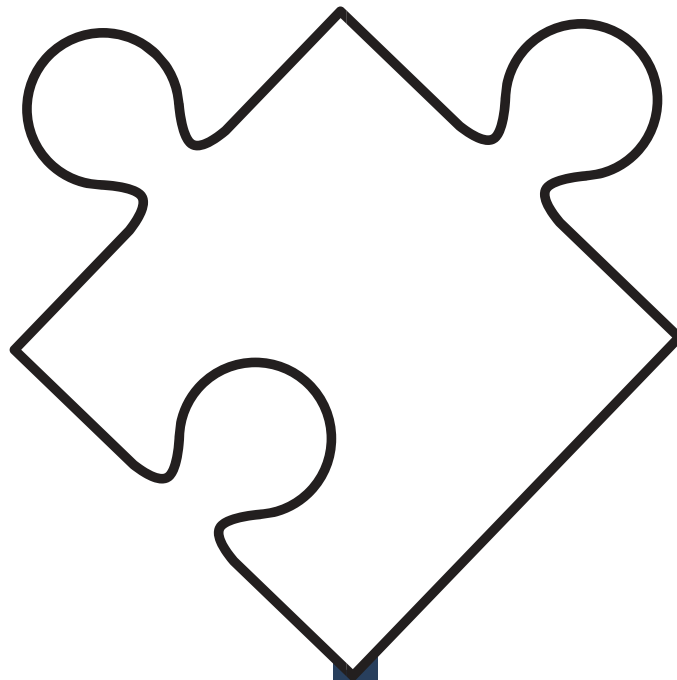


Proceedings of the 4th Student- STAFF Research Conference 2020



**UNIVERSITY OF
WESTMINSTER** 

**College of Design Creative and
Digital Industries**

**School of Computer Science and
Engineering**

**Proceedings of the
4th Student-STAFF Research Conference 2020
School of Computer Science and Engineering
SSRC2020**

27 October 2020
University of Westminster
United Kingdom

Editor:
Alexander Bolotov

Digital Object Identifier: <https://doi.org/10.34737/v2x14>

**UNIVERSITY OF
WESTMINSTER** 

© 2020 for the individual papers by the papers' authors. Reproduction (electronically or by other means) of all or part of this technical report is permitted for educational or research purposes only, on condition that (i) this copyright notice is included, (ii) proper attribution to the editor(s) or author(s) is made, (iii) no commercial gain is involved, and (iv) the document is reproduced without any alteration whatsoever. Re-publication of material in this technical report requires permission by the copyright owners.

SSRC2020 Organising Committee:

Alexander Bolotov, Chair

Session Chairs

Thierry Chausalet

Gabriele Pierantoni

Sophie Triantaphillidou

Natalia Yerashenia

Tamas Kiss

Preface

This volume contains the proceedings of the 4th Student-STAFF Research Conference of the School of Computer Science and Engineering (SSRC2020). This is a traditional, annual forum which brings together, for an one-day intensive programme, established and young researchers from different areas of research, doctoral researchers, postgraduate and undergraduate alumni, and covers both traditional and emerging topics, disseminates achieved results or work in progress. During informal discussions at conference sessions, the attendees share their research findings with an open audience of academics, doctoral, postgraduate and undergraduate students. The SSRC2020 was held on-line. The specifics of this year's conference was the participation of alumni from the Informatics Institute of Technology (IIT Sri Lanka) and Westminster International University in Tashkent (WIUT, Uzbekistan). The event met great interest - it had more than 200 on-line participants, with one session accommodating the audience of 156! The presenters whether they are established researchers or just at the start of their career, not only share their work but also gain invaluable feedback during the conference sessions.

Twenty one abstracts of the Proceedings contributed by the speakers at the SSRC2020 are assembled in order of their presentation at the conference. The abstracts cover a wide spectre of topics including the development of on-line knowledge and learning repositories, data analysis, applications of machine learning in fraud detection, bankruptcy prediction, patients mortality, image synthesis, graph DB, image analysis for medical diagnostics, mobile app developments, user experience design, wide area networking, adaptive agent algorithms, plagiarism detection, process mining techniques for behavioural patterns, data mining for reablement, Cloud Computing, Networking and linguistic profiling.

I would like to thank the members of the Organising Committee for their advice, assistance and chairing the sessions. I would also like to thank Kamalini Sivagurunathan and Kumudini Sarathchandra for their efforts to network with IIT Sri Lanka which significantly increased SSRC2020 audience.

London
December 2020

Alexander Bolotov

Contributed Papers

SMARTEST – Stressless and Interactive Learning to maximize engagement	1
<i>Alexander Bolotov, David Chan You Fee, Gabriele Pierantoni and Natalia Yerashenia</i>	
Automobile Insurance Fraud Detection using Predictive Algorithms	2
<i>Nichalini Kandasamy</i>	
Using Process Mining to map Students’ Learning Patterns	3
<i>Mike Jason Wigger</i>	
Semantic Data Pre-Processing for Machine Learning Based Bankruptcy Prediction Computational Model	4
<i>Natalia Yerashenia, Alexander Bolotov, David Chan and Gabriele Pierantoni</i>	
A New Machine Learning Modelling Approach for Patients’ Mortality Prediction in Hospital Intensive Care Unit	5
<i>Nodira Nazyrova, Mahmoud Aldraimli, Abdumalik Djumanov, Ikboljon Sobirov, Thierry J. and Chaus-salet5</i>	
An Analysis of Perceptions to Railway Line High-Speed Two: an Investigation Using Social Media and other Online Sources	7
<i>Kathryn Liddle</i>	
Image Analysis and Feature Extraction of Kato-Katz Images for Neglected Tropical Diseases Diagnosis	8
<i>Farah Sinada, Saumya Kareem Reni and Izzet Kale</i>	
SugaCheck	9
<i>Amanda Shehni Mansoori</i>	
Manifold Learning of Latent Space Vectors in GAN for Image Synthesis	10
<i>Mohamed Ayob</i>	
AutoSkreen: Automated Skin Cancer Screening using Optimized CNN	11
<i>Minoj Selvarasa</i>	
Dynamic and digital style guide system for Agile environments	12
<i>Naveen Kavishma Dolamullage Jayasinghe</i>	
Evaluation of Routing Protocols and WAN Configuration	13
<i>Kyriaki-Maria Valavanidou</i>	
Adaptive Agent for FPS Games using Reinforcement Learning	14
<i>S.A Didula Kasunka Manura</i>	
GhostBuster – A writing style detection system to identify ghost-written plagiarism	15
<i>Namila Dammalage</i>	
InfoChurn- A Customer Retention Engine with Churn Prediction for Motor Insurance in Sri Lanka	16
<i>A.R.A.Thilini Prarthana Gimhani and Kumudini Sarathchandra</i>	
GContactless Loyalty	17
<i>Shouyi Cui</i>	
Evaluating the effectiveness of Reablement as a service	18
<i>Bersanta Coku</i>	
Cloud-Based Mobile Access Control	19
<i>Danislav Lyubenov</i>	
SinTM - LDA and RAKE based Topic Modelling for Sinhala	20
<i>Dinushika Rathnayake and Saman Hettiarachchi</i>	
Person Verification using Linguistic Profiling for Continuous Authentication Based On ML	21
<i>Murtaza Anverali</i>	
Using Artificial Intelligence Technology and Machine Learning techniques to create a Bankruptcy Calculator	22
<i>Sravan Manda</i>	

SMARTEST – Stressless and Interactive Learning to maximize engagement*

Alexander Bolotov¹, David Chan You Fee², Gabriele Pierantoni³ and Natalia Yerashenia⁴

University of Westminster, London, United Kingdom,

¹a.bolotov@westminster.ac.uk

²chanyod1@westminster.ac.uk

³g.pierantoni@westminster.ac.uk

⁴w1578366@my.westminster.ac.uk

Keywords: on-line learning platforms, personalised learning, graph based knowledge and learning repository, conceptual maps, learning interfaces.

SMARTEST is an online learning tool (<https://smarrestknowledge.org/>), supported on desktop and mobile, which utilises graphs to conceptually map out learning processes and present them visually and interactively to learners. The platform is content rich – expressing learning and knowledge structures in a graphical format, it allows to break down the problem into individual steps and represent visually as a graph. Seeing various flows, a student can adopt the one which suits them better, and choose a flow which they feel is best.

SMARTEST uses a custom editor which offers an intuitive interface to construct graphs, and options for adding metadata to nodes and relationships like a name, description, links and tags. A messaging system acts as a channel of support for students needed assistance. “Graph linking” allows to develop small graphs that are linked to many other graphs containing more information. SMARTEST also supports collaboration, where multiple users can work together on the same graph and different languages for graphs, which is a valuable feature for users in Tashkent. A PDF export feature, which abstracts a graph’s contents into a single document, allows learning content to be viewed offline.

In the future, we would like to implement a SMARTEST API, multi-language support for the general interface and integration with the Blackboard VLE.

Automobile Insurance Fraud Detection using Predictive Algorithms

Nichalini Kandasamy

University of Westminster, London, United Kingdom

nicha.kandasamy@gmail.com

Keywords: Machine Learning, Insurance Fraud, Predictive Algorithms, Classification Methods

Insurance fraud is a rising problem affecting businesses and members. Through the years, several works have been led to detect fraudulent claims at an early stage and before paying out the claims. However, the methods have not been proven to be sufficient enough to tackle the problem.

Machine Learning techniques have considerably expanded and have the ability to solve complex problems. Multiple Machine Learning methods have been studied and applied in the detection of fraudulent claims. Nevertheless, not all available Machine Learning methods have been examined. Certain algorithms remain a mystery in the field.

In the following project, five algorithms will be formed, analysed and compared : Naïve Bayes, Random Forest, Deep Learning, Support Vector Machines and Logistic Regression. Out of the five algorithms, Naïve Bayes and SVM are the most frequently employed algorithms for fraud detection. The data used for the development of the models is a sample data of a car insurance company's database, located in the USA. The data has been pre-processed and cleansed before the creation of the models.

The comparison of the models ,after building them on the available data, outlined that the Random Forest and Deep Learning models were effective and precise in the training and learning procedures. Hence, leaving us questioning, the effectiveness of the existing techniques.

Using Process Mining to map Students' Learning Patterns

Mike Jason Wigger

University of Westminster, London, United Kingdom

Mike.Wigger17@gmail.com

Keywords: Process Mining, e-Learning, Behavioural patterns, ProM, Disco Fluxicon.

Considerable research has been carried out on the analysis of data resulting from students' interactions through e-learning. Several aim at discovering patterns in the behaviour of students through Process Mining techniques. The research project carried out sought to analyse the data collected from Blackboard to address two questions. In the first phase, the researcher wants to determine whether there are any variations in the behaviour of students using an e-learning system. If so, it is necessary to identify then whether a group of highperforming learners follow alternative patterns compared to others. To address these questions, several experiments and analyses will be conducted to discover process models illustrating the students' activity on Blackboard. This step will involve the use and comparison of various tools and techniques, such as ProM and Disco Fluxicon as well as the process discovery algorithms like the Alpha algorithm, the Fuzzy Miner, and the Heuristics Miner. Ultimately, this methodology will prove that there are indeed significant dissimilarities in the behaviour of students through distance learning. Students who have completed the programming module with distinction, have shown a stronger activity and a more assiduous behaviour than the rest of the class.

Semantic Data Pre-Processing for Machine Learning Based Bankruptcy Prediction Computational Model

University of Westminster, London, United Kingdom

Natalia Yerashenia¹ Alexander Bolotov² David Chan³ Gabriele Pierantoni⁴

University of Westminster, London, United Kingdom

¹w1578366@my.westminster.ac.uk

²A.Bolotov@westminster.ac.uk

³davidchan125@hotmail.co.uk

⁴G.Pierantoni@westminster.ac.uk

Keywords: semantic data analysis, graph database, ontology, financial analysis, financial ratios, bankruptcy prediction, computational model, neural network, Protégé, Neo4j, Python

The project studies a Bankruptcy Prediction Computational Model – a comprehensive methodology of evaluating a company’s bankruptcy level, which combines storing, structuring and pre-processing of raw financial data using semantic methods with machine learning analysis techniques. The main goal of our research is to develop data pre-processing techniques where ontologies play a central role. We show how ontologies are used to extract and integrate information from different sources, prepare data for further processing, and enable communication in natural language. Our Ontology of Bankruptcy Prediction which provides a conceptual framework for a company’s financial analysis, is built in the widely Protégé environment. An OBP Ontology can be effectively described with a Graph database. A Graph database expands the capabilities of traditional databases by tackling the interconnected nature of economic data and providing graph-based structures to store information, allowing the effective selection of the most relevant input features for the machine learning algorithm. To create and manage the BPCM Graph database, we use the Neo4j environment and Neo4j query language, Cypher, to perform feature selection of the structured data. Selected key features are used for the supervised Neural Network with a Sigmoid activation function. The programming of this component is performed in Python.

A New Machine Learning Modelling Approach for Patients' Mortality Prediction in Hospital Intensive Care Unit

Nodira Nazyrova¹ Mahmoud Aldrainli² Abdumalik Djumanov³
Ikboljon Sobirov⁴ and Thierry J. Chausalet⁵

¹ Westminster International University in Tashkent, nnazyrova@wiut.uz

² University of Westminster, m.aldrainli@my.westminster.ac.uk

³ Westminster International University in Tashkent, adjumanov@wiut.uz

⁴ Westminster International University in Tashkent, i.sobirov@wiut.uz

⁵ University of Westminster, chausst@westminster.ac.uk

Keywords: Imbalanced Learning, Machine Learning, Mortality, Severity Scores, Classification, Sampling, Ensembles

Mortality prediction in a hospital Intensive Care Unit (ICU) is a challenge that must be addressed with high precision. Machine Learning (ML) is a powerful tool in predictive modelling but subject to the problem of class imbalance [1]. In our study, we tackle class imbalance with combining new features, data re-sampling [2], ensemble learning [3] and an appropriate selection of evaluation metrics. We built and evaluated 126 ML models to predict mortality in 48546 ICU admissions extracted from the Medical Information Mart for Intensive Care III (MIMIC-III) [4] repository.

In our study design, six mortality prediction datasets are extracted; five of which are legacy dataset sets while the remainder is our new constructed dataset. For our combined data models, our selection of features enhanced the prediction performances beyond those for the traditional legacy sets used in research. Our approach has a considerable impact on the classification; it resulted in an improvement in the mortality status prediction. For evaluation, we implement a comparative multi-stage evaluation filter for a binary classification to compare all models. Area Under Receiver Operator Characteristic curves [5] of the tested models range from 0.57 to 0.94. These encouraging results can guide further development of models to allow for more reliable ICU mortality predictions.

References

- 1 Rahman M, Davis D (2013) Addressing the Class Imbalance Problem in Medical Datasets. International Journal of Machine Learning and Computing, pp. 224-228
- 2 Chawla N, Bowyer K, Hall L, Kegelmeyer W (2002) SMOTE: Synthetic Minority Over-sampling Technique. Journal of Artificial Intelligence Research (JAIR) 16: 321-357
- 3 Wu Z, Lin W, Ji Y (2018) An Integrated Ensemble Learning Model for Imbalanced Fault Diagnostics and Prognostics. IEEE Access, vol. 6, pp. 8394-8402, doi: 10.1109/ACCESS.2018.2807121.

- 4 MIMIC-III, a freely accessible critical care database. Johnson AEW, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, Moody B, Szolovits P, Celi LA, and Mark RG. *Scientific Data* (2016). DOI: 10.1038/sdata.2016.35. Available at: <http://www.nature.com/articles/sdata201635>
- 5 Andrew PB (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30: 1145–1159. DOI:[https://doi.org/10.1016/S0031-3203\(96\)00142-2](https://doi.org/10.1016/S0031-3203(96)00142-2)

An Analysis of Perceptions to Railway Line High-Speed Two: an Investigation Using Social Media and other Online Sources

Kathryn Liddle
University of Westminster, London, United Kingdom,
w1712301@my.westminster.ac.uk

Keywords: Social Media Analytics, Large Scale Projects, Public Perception.

Social media analytics (SMA) has shown to bring benefits such as increased brand awareness within many industries, for example entertainment. However, large-scale infrastructure (LSI) projects have been slow to embrace SMA.

Therefore, to explore whether SMA can provide similar benefits to LSI projects, this study investigates the public perception to High-Speed Two (HS2) using Twitter data from 2018 to 2020.

Firstly, to investigate the public perception towards HS2 and the reasons for this, sentiment analysis and topic modelling were executed. Analysis of engagement was also executed to interpret patterns of retweets. Secondly, to investigate the effect of news articles on public perception, news articles have been analysed and compared to changes in sentiment and topics of discussion on social media.

The findings show that the public perception towards HS2 is mixed, however, evidence suggests it became more negative over time. The most frequent topics discussed were 'Transport' and 'Rail Network', however 'Cost' and 'Environment' were shown to be important topics due to high engagement and negative sentiment. It was observed that news articles do have somewhat of an impact on the sentiment and topics discussed.

This study demonstrates that valuable insights can be extracted through SMA for LSI projects.

Image Analysis and Feature Extraction of Kato-Katz Images for Neglected Tropical Diseases Diagnosis

Farah Sinada¹ Saumya Kareem Reni² Izzet Kale³

University of Westminster, London, United Kingdom

¹farahmakram@gmail.com

²S.Reni@westminster.ac.uk

³kalei@westminster.ac.uk

Keywords: Neglected tropical diseases, Helminthiasis, kato-Katz Images, Morphological Image processing, Feature extraction, MSER

This paper provides an insight into a pioneering interdisciplinary research, amalgamating engineering principles and clinical diagnostic procedures. The research is underway to design and develop an automated parasite diagnostic tool, using machine learning based image processing operations. The preliminary findings of the work in which feature extraction algorithms are applied on clinical images as a base to a neural network model for diagnosis is described in this paper. Maximally Stable Extremal Regions (MSER) was used as a primary feature extractor in the detection of the parasitic eggs. Statistical information regarding the eggs was then collected and compared with background regions in the development of a search algorithm. Using the collected data, a range of values was obtained that could separate the eggs from the background of the images. All the images used are from Kato-katz slides.

SugaCheck

Amanda Shehni Mansoori

University of Westminster, London, United Kingdom,
w1594861@my.westminster.ac.uk

Keywords: Android, Mobile, Application, Diabetes.

Our goal was to create and develop an application that would help diabetic individuals with their everyday tasks. The initial steps that were taken included an intensive research on diabetes, exploration of already existing applications and understanding the most needed functions the application would need to offer. The application allows the user to insert their own insulin ratios, set their preferred blood glucose ranges, specify when their blood sugar is deemed too high or low and set up notifications and reminders. The user can also add entries with information such as their blood sugar level, the amount of carbohydrates they have consumed and the insulin they have taken. The application makes suggestions on how much insulin to take based on the details entered, and is supported by a food database, allowing the user to search for food items or scan an item using the barcode scanner to get the nutritional values. The application also offers statistics, displaying insulin, blood glucose and carbohydrate averages. Living with diabetes can be very difficult and having this application can provide the help and support needed to keep track of all the data in a seamless and effortless way.

Manifold Learning of Latent Space Vectors in GAN for Image Synthesis.

Mohamed Ayoob

University of Westminster, London, United Kingdom
w1654551@my.westminster.ac.uk

Keywords: Image Generative Model; Generative Adversarial Networks; Latent Space Vectors; Mathematical optimization; Image synthesis, Face recognition systems.

Generative Adversarial Networks have attained remarkable results in image augmentation and generation. GAN models consist of Latent space vectors. Mathematically they are a hypersphere consisting of many dimensions, containing the learnt representation. However, various GAN variants use different techniques to model latent space vectors. This dissertation dissects the latent space vectors of state-of-the-arts GAN variants, for image synthesis. We also compare and contrast optimization methods of latent space vectors in existing literature. In this dissertation, we undertake a comparative analysis of latent space vectors and optimization methods of the popular image synthesizing GAN models.

Based on a review of mathematical methods we choose manifold learning techniques to cluster the latent space. Results of clustering allowed us to determine the sparse and dense features of Euclidean image distribution.

The research on the image augmentation techniques allowed the authors to make a deployable facial recognition system utilizing advanced image models. The authors had to review existing work on face recognition such as FaceNet model to make the system. The authors used concurrent programming principles to optimize the model to train in reasonable amounts of time. This will be useful with pandemics such as the COVID-19 they are a non-invasive (touchless) authentication system.

AutoSkreenr: Automated Skin Cancer Screening using Optimized CNN

Minoj Selvarasa

Informatics Institute of Technology, Sri Lanka

minoj.2016131@iit.ac.lk

Keywords: Skin Cancer Screening, Image Classification, CNN, Optimisation Algorithms

Cancer is one of the deadliest diseases known to mankind. Skin cancer is a common form of cancer that is known to commonly affect people of lighter skin. Survival rates of those diagnosed at an early stage is much higher than in the later stage. Statistics show that the number of skin cancer cases have been on the rise in the past few years in western countries while the number of dermatologists have been constant.

This calls for a solution that can aid dermatologists in their screening process. This research involves building a CNN based Image Classification tool that can help dermatologists in identifying the exact type of skin cancer. The research involves the exploration of Optimization algorithms to tune the CNN and comparing experiments to identify the most optimal experiment.

Dynamic and digital style guide system for Agile environments

Naveen Kavishma Dolamullage Jayasinghe

Informatics Institute of Technology, Sri Lanka in collaboration with University of Westminster,
London, United Kingdom, kavishmaj@gmail.com

Keywords: User experience design, Style guide, Agile, Design thinking, User centered design, User interface design.

User experience (UX) is referred to as a person's perceptions resulting from the use of a product, system or service. It enables the application of empathy in web and software design to facilitate a better product experience. Despite its benefits, several studies have concluded that it has become a challenge for both Agile and UX to co-exist in a single environment. This has forced Agile to make compromises in UX design, which has led to the absence of design artefacts, such as style guides, that documents the design rationales of a system. The goal of this project is to evaluate the use of style guides and the application of UX within Agile software companies in Sri Lanka. A quantitative and qualitative study was conducted locally and it indicates that UX practitioners fail to enforce a thorough UX process in Agile teams due to various challenges, including limited resource allocation for design and research. Moreover, designers continue to use static style guide documents that are often neglected because they are too detailed and time-consuming to create, maintain and follow. Thus, a viable system was developed to support the creation of dynamic style guide documents for Agile software companies in Sri Lanka.

Evaluation of Routing Protocols and WAN Configuration

Kyriaki-Maria Valavanidou

University of Westminster, London, United Kingdom,
w1707181@my.westminster.ac.uk

Keywords: QoS, WAN, Wide Area Network, heterogeneous network, network, routing, protocol, simulation, analysis, EIGRP, IGRP, RIP, BGP, OSPF, IS-IS, security, Riverbed

Over the years, the progress of technology has brought forth many necessities and propositions, especially over networking subjects. The suggestion of multiple routing protocols is one of them. In this report, the findings of thorough study are presented in order to assess the efficiency of several routing protocols over a heterogeneous network environment; specifically, a Wide Area Network (WAN). There is also research conducted into the depths of a WAN, such as other protocols necessary for setting up a WAN, security issues discovered over time, as well as what an effective WAN would be consisted by.

Adaptive Agent for FPS Games using Reinforcement Learning

S.A Didula Kasunka Manura

University of Westminster, London, United Kingdom
Didula.2015351@iit.ac.lk

Keywords: Reinforcement learning, Temporal Difference learning, Q-learning, Sarsa learning, Epsilon greedy, Adaptive agent, First person shooter games, Pogamut.

Artificial game agent provides the adaptive functionality, to minimize the need of implementing behavior-based agent according to the environment, causing to reduce complexity of an agent implementation. Due to immense benefits provided by machine learning, artificial agents are implemented using these technologies. The common goal of these agents is to provide the entertainment to the end-users. To acquire the goal for an adaptive agent, various functionalities like dynamic reward system for increasing efficiency in learning process and dynamic weapon preference system for increasing the winning rate of gun fights are required. Some systems communicate with online algorithms such as Q-learning algorithms, and some systems communicate with offline algorithms, in order to achieve the ultimate goal of learning process. The only difference in such algorithms, is the allocation of reward to an action based on time.

Focusing on artificial agents, this project addresses these concerns through an adaptive agent, for first-person shooter game in deathmatch mode.

Dynamic information obtaining through the different states of the agent can be utilized in order to integrate a dynamic reward system. In the same way, information obtained from the weapons during combat states can be utilized to create dynamic weapon preference system based on damage.

GhostBuster – A writing style detection system to identify ghost-written plagiarism

Namila Dammalage

Informatics Institute of Technology Sri Lanka affiliated with University of Westminster, London, United Kingdom, dnamilal@gmail.com

Keywords: plagiarism detector, ghost-writing, writing style.

With a rapid increase in student populations across universities, the competition among students is extremely prominent. Gone are the days where a simple copy-paste off the internet will give students an edge over their colleagues. This is mainly due to the superior level of plagiarism detectors available in the market today. The dilemma arises when these world class plagiarism detectors are unable to detect when a specific report has been outsourced because there are no direct indicators available. Due to this reason a new market has opened for people that are experts in various fields where they complete these difficult assignments for the students for a price called “ghost-writers”.

GhostBuster would provide an in-depth analysis of the text characteristics of a known report of the student and the report that is under suspicion of ghost-written plagiarism. It will be able to provide quantitative evidence on how similar the 2 reports are thereby providing the markers of reports with enough evidence to apprehend the students that resort to ghost-written plagiarism. The solution will further provide graphical analysis of all the deductions that are made to enable users to understand with ease.

InfoChurn- A Customer Retention Engine with Churn Prediction for Motor Insurance in Sri Lanka

A.R.A.Thilini Prarthana Gimhani¹ Kumudini Sarathchandra ²

¹ University of Westminster, London, United Kingdom, thilini.2016501@iit.ac.lk

² Informatics Institute of Technology, Sri Lanka, kumudini.s@iit.ac.lk

Keywords: churn prediction, data mining, insurance industry

Companies are striving to explore new ways to retain their clientele to sustain their business growth due to economic volatility and heated competition in the insurance industry around the world. Retaining customers is more cost-effective than acquiring new customers thus the customer churn, a critically addressed topic in both industry and academia. Despite being a globally addressed problem, the Sri Lankan insurance industry has lagged in leveraging technology-based solutions to outsmart churn. Following the case study methodology, this study was aimed at analysing the churn behaviour of motor insurance customers of Ceylinco Insurance PLC.

The broader discipline of data mining using machine learning techniques was critically screened in deriving a churn prediction solution using a real-life data set containing 478,157 customer records. Data preparation included scaling data, handling class imbalance with SMOTENN technique while the most important features were selected using Pearson's correlation. From the classification algorithms, the Gradient Boosting classifier derived the highest F1 score of 91.95%. From the regression algorithms, Random Forest Regressor performed the best. The results are aimed at maximizing profits by classifying customers to introduce targeted marketing for policy renewals instead of wasting their time, effort and money trying to retain their entire clientele.

Contactless Loyalty

Shouyi Cui

University of Westminster, London, United Kingdom, w1618594@my.westminster.ac.uk

Keywords: NFC, Contactless, Loyalty, Web Application, Mobile, Smartphone.

This project aims to resolve the problem of the physical loyalty cards either made of paper or plastic. A research about the alternative digital solution is carried out to define what can be developed as a new service not yet in the market. Based upon a statistical approach and a real case business scenario a solution is provided with the Web-NFC API, which is different from the current mobile applications. The solution takes into account the budgets and the different limitations of small brands. The target audience of the service can be considered as the retailers in the small shops because they would make available the service for the customers. The feedback received from the testers describe a higher level of engagement between the retailer and the customer. At the end of this project, the software developed gives the opportunity for the customer to engage with the retailer. From the retailer point of the view, the web application implemented for this beta version does not meet all the requirements although it gives the essentials functionalities.

Evaluating the effectiveness of Reablement as a service

Bersanta Coku

University of Westminster, London, United Kingdom,
w1747624@my.westminster.ac.uk

Keywords: Reablement, LBHF, Data mining, classifiers

Many Reablement programmes are provided as part of the assistance following a hospital stay. Reablement provides to individuals who need support, to regain their autonomy or improve their quality of life.

The goal of this research is to analyse how productive the Reablement as a service is in regard to costs and also to identify any trends by comparing the patients who received, with the patients who did not receive Reablement. Data are provided by the Council of the London Borough of Hammersmith and Fulham for three years from 2018-2020. Classification techniques are one of the most used data mining techniques in the healthcare area. In this Research three different types of classifiers Random Forest, Decision Tree and Support Vector Machine are used to create a model to predict the target class which is, the likeliness of a patient to have Reablement as a service. Evaluating the model based on the prediction accuracy measurements, the Support Vector Machine could predict accurately 80% of the patients.

Cloud-Based Mobile Access Control

Danislav Lyubenov

University of Westminster, London, United Kingdom
w1592922@my.westminster.ac.uk

Keywords: Cloud Computing, Mobile Access Control, Near Field Communication

This project explores a big innovation that Cloud computing has contributed to individuals and businesses alike. By using public Cloud providers, buying and configuring network infrastructure will become a thing of the past. Instead, this project uses Cloud resources in the form of databases, file storage and image processing to create an Android application which integrates Near Field Communication technology to securely unlock hotel rooms and automate the process of guest validation.

SinTM - LDA and RAKE based Topic Modelling for Sinhala

Dinushika Rathnayake¹ Saman Hettiarachchi²

¹ Department of Computing, Informatics Institution of Technology, Ramakrishna Road, Colombo6, Sri Lanka, dinushika.r917@gmail.com

² Department of Computing, Informatics Institution of Technology, Ramakrishna Road, Colombo6, Sri Lanka, saman.h@iit.ac.lk

Keywords: Natural Language Processing, Topic Modelling, Latent Dirichlet Allocation, Rapid Automatic Keyword Extraction, Sinhala

Sinhala is the native language in Sri Lanka and primarily spoken by Sinhalese. Since the effects of technology improved the Sinhala text usage on the web, it's become hard to work with large amounts of textual information. Topic modelling is a Machine Learning(ML) technique used in Natural Language Processing(NLP) to retrieve abstract topics from a collection of documents. This technique is highly important to discover hidden and useful information from huge unstructured and heterogeneous text data.

This paper presents a novel approach called SinTM to analyse a single Sinhala text document by combining topic modelling and keyword extraction techniques. The primary contribution of this research is the experiment of topic modelling task for the Sinhala language. As a second contribution, this paper presents a novel hybrid approach called SinTM for Sinhala topic modelling. Also, the results were benchmarked with well-known topic modelling algorithm LDA as a third contribution.

The SinTM was tested with prominent topic model evaluation matrices likelihood, r-squared, perplexity and coherence. The web user interface comes with the SinTM system provides a more controllable parameter tuning and easy understandable graph-based view to the user and wealthy in terms of comparing the novel model against the well-known LDA model.

Person Verification using Linguistic Profiling for Continuous Authentication Based On ML

Murtaza Anverali

Informatics Institute of Technology, Sri Lanka

murtaza.2018594@iit.ac.lk

Keywords: Person Verification, Linguistic Profiling, Support Vector Machine, Multi-Layer Perception

The exponential increase in the usage of mobile devices has led to the rise of continuous authentication mechanisms. Continuous authentication involves constantly verifying a person in the device background. Person verification entails verifying a person uniquely using their specific traits. Various advancements have been made to carry out this type of verification including usage of linguistic profiling. Linguistic profiling involves using the user's stylometry for verification. In this study, while exploring person verification using linguistic profiling, a variety of supervised machine learning techniques were used.

A platform was developed by combining linguistic profiling with dynamic feature extraction and per chat model building. The system accepts WhatsApp chats and the algorithms used include Support Vector Machines (SVMs), Multi-Layer Perceptron (MLPs), and Ensemble algorithms. Training and testing were carried out on 146 WhatsApp chats, each having an average of 216 messages. The accuracy obtained from the platform includes a False Acceptance Rate (FAR) and False Rejection Rate (FRR) of 75.81%, 8.62% 15.5% respectively. Dynamic feature extraction was used to achieve this. This involves identifying the most common words, emojis, word extensions, etc. used. This enables the creation of a linguistic profile for the user's chat. Once models are trained, chat messages can be used to predict if the message is from an imposter or the actual user. The platform can also be leveraged by experts in forensic and cyber-crime domains.

Using Artificial Intelligence Technology and Machine Learning techniques to create a Bankruptcy Calculator

Sravan Manda

University of Westminster, London, United Kingdom

sravankml@gmail.com

Keywords: bankruptcy prediction, machine learning, graph DB

The problem to predict bankruptcy and falling into a financial risk is very difficult to expect by many businesses due to lack of such accurate prediction models. The aim of this project is to investigate and work with as many different aspects of a business along with its data sets to create an algorithmic approach which provides an accurate enough value that could determine the probability of bankruptcy for the organisation. My proposal to solve this issue is to use many of the current and advanced tools such as Neo4j and MATLAB along with machine learning techniques which are thought to be an enabler for data mining to analyse existing bankrupt and non-bankrupt business data. This project develops a user interface comprising the factors obtained from the data mining to calculate a reasonable prediction of falling into bankruptcy.

Demo Link: <https://youtu.be/wGDlydyR1ws>