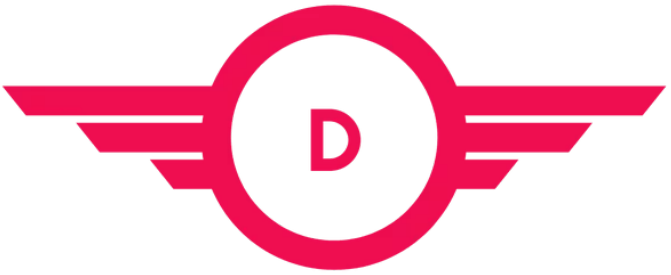
DSFSI Seminars



Data Science for Social Impact

Dynamic event detection on social media data streams. [1 March 2019]

Speaker: Jessica Nemasisi

Abstract: Event detection has attracted a lot of significant amount of research interest in social media analysis. However, most existing approaches from related literature focus on event detection applied on batch data.In this study we present an algorithm that efficiently detects events on text streams. The algorithm addresses the challenges of prior algorithms in that it does not need to constantly reuse all prior text documents, it further adaptively resolves the temporal scale needed to connect prior documents and tries to reduce the computational overhead by using a hashing approach. We apply our algorithm to the challenge of dynamic event detection on social media data.
Bio: Jessica Nemasisi is a Data Science masters student at the University of Witwatersrand and based at the CSIR. Jessica holds a computer science honours degree from the University of Venda. She was a data science intern at the CSIR (Modelling and Digital Science) from September 2015 until September 2016. Jessica hailds from Venda, Limpopo, in a rural area called Tshidimbini. One of her life goals is to introduce more rural girls like herself to all the cool things one can do as a Data Scientist.

Bayesian Inference in Neural Networks via Hamiltonian Monte Carlo [29 March 2019]

Speaker: Rendani Mbuvha

Bio: Rendani is a PhD Candidate in Artificial Intelligence at the University of Johannesburg and a lecturer in the School of Actuarial Science and Statistics at the University of Witwatersrand. He is a fellow of the Actuarial Society of South Africa and a Chartered Enterprise Risk Actuary. He holds a Bsc Honours in Actuarial Science from the University of Cape Town and an Msc in Machine Learning from KTH, Royal Institute of Technology in Sweden.

Authorship Identification in Social Media: Deep Learning Approach Presenter [29 March 2019]

Speaker: Abiodun Modupe

Abstract: To identify and track cybercriminals in social media platforms, investigators employ carefully engineered textual features along with traditional text mining techniques to capture criminal signatures. This process is usually manual, time-consuming and highly dependant on the engineered textual features which may result in retrieved pieces of evidences not trustworthy for pre-trial investigations. In this work, we aim to enhance this

process. We employ a multi-channel convolutional neural network over textual characters to extract linguistic features.
Bio: Abiodun is a PhD Candidate at the University of the Witwatersrand and a visiting PhD student at the CSIR Modelling and Digital Science.

Machine Learning for Environmental Conservation [26 April 2019]

Speaker: Ciira Maina, Kenya

Abstract: In this talk we explore the use of machine learning to monitor ecosystem biodiversity. In particular. we report on an ongoing experiment in the Mt Kenya ecosystem in Central Kenya employing acoustic recordings to determine bird biodiversity which serves as a proxy for biodiversity.
Bio: I am a lecturer at Dedan Kimathi University of Technology in Nyeri, Kenya where I also conduct research in a number of areas including bioacoustics, IoT, machine learning and data science. I am also a founder member of Data Science Africa. Prior to joining DeKUT in 2013, I was a postdoctoral researcher at the University of Sheffield between 2011 and 2013, a PhD student at Drexel University in Philadelphia, USA between 2007 and 2011 and a BSc Student at the University of Nairobi between 2002 and 2007.

AI Accelerated Mapping for Humanitarian and National Security Missions [22 May 2019]

Speaker: Dalton Lunga

Abstract: Geographic maps of human settlements are a critical input for tackling various real-world challenges ranging from poverty reduction to disaster relief. However, due to high costs, there is a severe shortage of such products in many places, especially in the developing world. Researchers at Oak Ridge National Laboratory (ORNL) are developing large scale automated feature extraction methods that are reliable and can efficiently map man-made structures to the level of individual buildings and solar panels using satellite or aerial imagery. The techniques have mapped large amounts of human settlements that were not accounted for in existing digital maps. The derived maps are providing core and substantial support for humanitarian missions by different organizations.
Bio: Dalton Lunga is a Staff Scientist in Artificial Intelligence, Machine Learning and Geographic Data Sciences at the Oak Ridge National Laboratory, US. He currently leads general purpose machine learning research in various areas including large-scale object detection for land use and land cover mapping, advanced workflows for ML deployment in high performance computing environments, large scale workflows for image search and retrieval, and representation learning with large varying density data. Prior to joining ORNL, Dalton was a senior researcher at the Council for Scientific and Industrial Research (CSIR), South Africa, working on machine learning for natural language processing, manifold

learning, and interactive visual analytics systems. While at CSIR he founded the Data Science for Impact and Decision Enablement (DSIDE) program in 2013. He serves on multiple professional organizations including his role as program chair for Artificial Intelligence and Geographic Knowledge Discovery (GeoAI Workshops), and editorial review board for journal frontier in artificial intelligence. He received his MS. and Ph.D. in Electrical and Computer Engineering from Purdue University, West Lafayette as a Fulbright scholar.

Using satellite images and computer vision to study the evolution and effects of spatial apartheid in South Africa. [2 August 2019]

Speaker: Raesetje Sefala

Abstract: This talk will explore the idea of using computer vision to classify neighbourhoods on satellite images from South Africa. We will also explore the relationship between the spatial and socioeconomic makeup of neighbourhoods(groups of houses; a community) in South Africa.

Bio: Raesetje Sefala is a computer science master's student at Wits University in South Africa. Her research interests include applying machine learning to help solve problems experienced by countries in the developing world. Her master's work is about using satellite images and computer vision to study the effects and evolution of spatial apartheid in South Africa and she is a recipient of the Data Science for Social Good fellowship at the University of Chicago.

A Digital Forensic and Machine Learning Approach for Discovering Behavioural Patterns of Online Sexual Predators [2 August 2019]

Speaker: Hombakazi Ngejane

Abstract: The digital ecosystem makes cyber-related sexual exploitation of minors a border-less crime. This requires effective tools to identify sexual predators who groom children using social media platforms. In this work, we investigate the use of Machine Learning (ML) and Natural Language Processing methods, guided by the Integrated Digital Forensic Process Model, to aid the digital forensic investigation. We propose interpretable ML as a tool that can identify the behaviour of potential sexual predators on chat platforms. We use ML models such as logistic regression, XGBoost, multilayer perceptron and bidirectional long short-term memory recurrent neural network, with varying interpretability power, to unearth the behavioural patterns. This behaviour can be used as a portfolio of evidence and also be used to confirm the theory of sexual predator behaviour.

Bio: Hombakazi is from rural areas of the Eastern Cape in a small town called Mount Frere. Studied Computer Science at WSU and completed Honours at UNISA. Currently working at CSIR Defense and Security as a Researcher and a MSc Computer Science student ath UP. Hombakazi has worked mainly on Data (profiling, reporting and visualization), Databases and a bit of software development. She developed an interest in Data Science and NLP through CSIR Data Science team previously based at the previous Modelling and Digital Sciences Unit.

The South African SDG Hub - Bridging the gap between research and policy. [6 September 2019]

Speakers: Willie Fourie, Hannes Strydom, Christopher Marais

Abstract: The Sustainable Development Goals (SDGs) are universal goals adopted by the United Nations to drive global development. The South African SDG Hub fosters evidence-informed policy making related to the SDGs by means of four workstreams: knowledge sharing, policy advice, dialogue promotion and capacity building. One of the Hub's most important tools is a repository that collects, tags and disseminates South African research on the SDGs. Extracting valuable information from this resource and interpreting it appropriately poses some technical challenges. Thus far we have experimented with Non-negative matrix factorization (NMF) and Latent Dirichlet allocation (LDA) topic modelling for text summarization. The aim is to commence classifying the research text of the repository according to the SDGs and their targets as an initial step of extract meaning and value from the dataset. The prospect of natural language processing through machine learning methods seem to allow for an efficient approach for tapping into this resource.

Mail and Guardian Data Desk [4 October 2019]

Speakers: Athandiwe Saba and Jacques Coetzee

Abstract: M&G Data Desk Projects at the intersection of journalism and Natural Language Processing.

Bio: **Athandiwe Saba** is a multi-award-winning investigative data journalist who is passionate about data, human interest issues, and good governance. She has worked for three of the biggest newspapers in South Africa and recently established the Mail & Guardian Data Desk which she now heads up. She has been internationally recognised for her work in data journalism by the Global Editors Network. She has studied at the University of South Africa and has studied finance at the Gordon Institute of Business Science through the Bloomberg Media Initiative. She was also awarded the News Corp Media Fellowship for African Journalists which allowed her to work at The Times (London) and The Wall Street Journal.

She is one of the authors of We are Going to Kill Each Other Today: The Marikana Story. Athandiwe is currently working on a large scale project about how just the justice system is, using statistical modelling with financial support through a grant by the Indigo Trust.

Bio: **Jacques Coetzee** is the Adamela Data Fellow at the Mail & Guardian, a position funded by the Indigo Trust. With a passion for discovery, he wears many hats: a data analyst and a journalist.

The process of creating satellite image ground-truth datasets for machine learning [19 June 2020]

Speaker: Raesetje Sefala
Sign Up: RSVP
Date: 19 June 2020
Time: 9AM-10 AM SAST

Abstract
This talk looks into the process of creating datasets in general for Machine Learning, some of the issues involved, tools for data labelling, evaluation processes and some of the challenges I went through in my own work while creating a dataset for labelling South African neighbourhoods on satellite images.

Speaker Bio
Raesetje is a Machine Learning Researcher who is currently a Computer Science Masters Student at Wits University, Johannesburg. Her research focuses on creating ground truth datasets and using machine learning to study spatial segregation in South Africa, post-Apartheid.

She is interested in building communities which aim to increase the capacity and quality of work, of underrepresented groups in AI. Raesetje has been involved in building and organizing events for communities such as Women in Computational Science Research and the Deep Learning IndabaX Pretoria.

She is mainly interested in using AI to solve problems experienced in developing countries; creating datasets for machine learning research and the discussions & creation/amendment of data privacy, ethics and accountability policies.

The seminars are hosted by the Data Science for Social Impact Research Group, in the Department of Computer Science at the University of Pretoria.

Using Machine Learning to Extract Key Meta-Data from Legal Text in Malawi [3 July 2020]

Speaker: Amelia Taylor
Sign Up: RSVP
Date: 3 July 2020
Time: 9AM-10 AM SAST

Abstract
Malawi faces a serious problem when it comes to law reporting and legal research. On one hand are the issues of accessibility and the availability and the scattered nature of the official reports. On the other hand are the challenges coming from the fact that the current document structure of Malawi legal text, e.g., court judgments, does not support a system of citation that makes it possible to link statutory law, case law and secondary law or to search by “legal terms” and their specific interpretations. This research tackles the specific problem of classifying court judgments disseminated by the High Court Library. The court judgments disseminated via the Malawi High Court Library are not classified according to useful categories, such as courts, topics of the law, statues they refer to. They do not have an index and the structure of the documents is not uniform. The internal structure of judgments impacts the efficiency of a search. The aim of this research was to develop a methodology for a semi-automatic classification of judgments disseminated by the High Court Library of the Malawi Judiciary with the purpose of enabling ‘intelligent searching’ within this body of knowledge. In this talk, we will discuss the processes involved in this research, challenges and opportunities.

Speaker Bio
Amelia graduated with a PhD in Mathematical Logic from Heriot-Watt University in 2006 where I was part of the ULTRA group. After that she worked as a research assistant on a project with Heriot-Watt University and the Royal Observatory in Edinburgh, aiming at developing an intelligent query language for astronomical data. From 2006 to 2013, Amelia also worked in finance in the City of London and Edinburgh - she built risk models for asset allocation and liability-driven investments. For the last 5 years, Amelia has been teaching programming and AI courses at the University of Malawi in the CIT and engineering department. Amelia also teaches research methodology and supervise MSc and PhD students. While my first interest in AI as an undergraduate was in the field of Natural Language Processing and intelligent query systems, she is interested in the other use of technology and AI for solving real-world problems. Amelia is always excited to learn and try new approaches and technologies such as virtual reality. Amelia is married and with3 children. Her husband grew up in Malawi. She loves travelling, people. culture and

encourages everyone to come and see the beauty of Malawi.
https://www.linkedin.com/in/amelia-taylor-phd-80938544/