

Monday, 26th November 2018

Arrival and Conference Registration --- Open for the whole day

Venue: Link area

Time	Venue: Room 301	Venue: Council Chambers	Venue: Thamsanqa (Auditorium)	Venue: Bram Fischer
	Workshop - Full Day Causal Inference and Propensity Score Analysis <i>Wei Pan</i>	Workshop – Full Day Bayesian Biostatistics Data Analysis <i>Emmanuel Lesaffree</i>	Workshop – Morning Getting Started with Research in Data Science: developing statistical insights via interactive data visualisations <i>Jim Ridgway</i>	Workshop – Morning Spatial statistics for health research <i>Louise Ryan</i>
08h30 – 10h00	Session 1	Session 1	Session 1	Session 1
10h30 – 11h00	Tea and coffee break	Tea and coffee break	Tea and coffee break	Tea and coffee break
11h00 – 12h30	Session 2	Session 2	Session 2	Session 2
12h30 – 13h30	Lunch	Lunch	Lunch	Lunch
			Workshop - Afternoon Creating and editing Statistical output in SAS® <i>André Zitzke</i>	Workshop – Afternoon Ensemble classification <i>Berthold Lausen</i>
13h30 – 15h00	Session 3	Session 3	Session 1	Session 1
15h00 – 15:30	Tea and coffee break	Tea and coffee break	Tea and coffee break	Tea and coffee break
15h30 – 17h00	Session 4	Session 4	Session 2	Session 2
17h00 –	SASA EC Meeting Venue: NB Pityana Room 02-56			
17h00 – 19h00	Registration Venue: Call centre			
17h00 – 20h00	Welcome Cocktail function Venue: Call centre			

Tuesday, 27th November 2018

07:00 -	Registration Venue: Link area				
	Morning tea and coffee				
08h30 – 10h30	Opening Ceremony Venue: Thamsanqa Kambule Auditorium				
	Welcoming:		Prof Danie Uys, President: SASA		
	Opening:		Prof Bhekile Mamba – Dean of the College of Science, Engineering and Technology, UNISA		
	Keynote address:		Prof Riaan de Jongh Professional Statisticians / Data Scientists: Who are they and how can we train them?		
	AWARDS Presentations:		SAS Awards for best honours projects Postgraduate paper competition Sichel Medal Thought leader award Fellowship and Honorary Members		
	Platinum Sponsor Address:		SAS: Mr André Zitzke		
10h35-11h10	Tea and coffee				
	Stream 1: Methods and Theory Venue: Council Chambers Chair: <i>Prof James Allison</i>	Stream 2: Bayesian Statistics Venue: Thamsanqa Kambule Chair: <i>Dr Gaetan Kabera</i>	Stream 3: Business and Industrial Statistics Venue: Room 2-57 Chair: <i>Dr Jean-Claude Malela Majika</i>	Stream 4: General and Applied Statistics Venue: Room 301 Chair: <i>Dr Jaco Visagie</i>	Stream 5: Young Statisticians Venue: Bram Fischer Chair: <i>Dr Njeri Wabiri</i>
11h10 -11h30	Distribution of the condition number from dual noncentral Wishart type matrices <i>Johan Ferreira</i>	Investigation comparing simulation methods for models with common constraints <i>Sean Van der Merwe</i>	Statistical modelling of Zimbabwe's international tourist arrivals using both symmetric and asymmetric volatility models. <i>Tendai Makoni</i>	The effects of mis-specification of the deterministic components in the Co-integration model <i>Sagaren Pillay</i>	The Classification of Galaxy Images using Convolutional Neural Networks <i>James Lloyd</i>

11h35 – 11h55	Probability models for predicting electrical system faults and estimating voltage quality loss by machine learning <i>Mphiliseni Nzuzo</i>	On Model Comparison: Application of Savage-Dickey Density Ratio to Bayes Factor <i>Olawale Akanbi</i>	Sequential rank Shiryaev-Roberts CUSUMs <i>Corli van Zyl</i>	Selecting Relevant Features for Multi-Label Classification of Emotions in Music <i>Trudie Sandrock</i>	Identifying Immunological Risk Factors for TB Progression Using a Hidden Markov Model <i>Miguel Rodo</i>
12h00 - 12h20	Probabilistic Hail Hazard and Risk Analysis <i>Ansie Smit</i>	Modelling extreme observations through an extended Topp-Leone distribution. <i>Andrehette Verster</i>	Analysis of Dependence Structure Between the Rand/USD exchange rate and the Gold/Platinum prices <i>Kajingulu Malandala</i>	Long-term peak electricity demand forecasting in South Africa: A quantile regression averaging approach <i>Norman Maswanganyi</i>	Visualisations for identifying missing data mechanisms <i>Johan Nienkemper-Swanepoel</i>
12h25 – 12h45	Testing for serial independence in vector autoregressive models <i>James Allison</i>	Bayesian Spatial-Temporal modeling of malaria risk in Nigeria <i>Ropo Ebenezer Ogunsakin</i>	Design of a side-sensitive double sampling X^- scheme for monitoring an unknown process mean <i>Jean-Claude Malela Majika</i>	On the use of Laplace's first and second laws of error in the save sample method <i>Jaco Visagie</i>	Bivariate threshold excess models with application to extreme high temperatures in Limpopo province of South Africa <i>Murendeni Maurel Nemukula</i>
12h50 – 14h00	Lunch				
12h55	MDAG Meeting (Venue: NB Pityana Room 02-56)				
	Stream 6: Methods and Theory Venue: Council Chambers Chair: <i>Prof Edmore Ranganai</i>	Stream 7: Bayesian Statistics Venue: Thamsanqa Kambule Chair: <i>Mr Allan Clark</i>	Stream 8: Statistics in Education Venue: Room 2-57 Chair: <i>Prof Delia North</i>	Stream 9: General and Applied Statistics Venue: Room 301 Chair: <i>Ms Jani Deyzel</i>	Stream 10: Young Statisticians Venue: Bram Fischer Chair: <i>Prof Principal Ndlovu</i>
14h00 – 14h20	Constant versus Covariate Dependent Threshold in the Peaks-Over Threshold Method <i>Richard Minkah</i>	Application of Extreme Value Theory (EVT) to Inflation Rate Targeting: The Ghanaian case. Occupancy models using logit link functions <i>Ezekiel Nortey</i>	Teaching Statistics for Social Change <i>Jim Ridgway</i>	Super rugby: modelling injury and illness <i>Charl Janse van Rensburg</i>	Advances and consequences of hot deck imputation <i>Nyiko Khoza</i>
14h25 - 14h45	Efficiency of empirical characteristic function estimators with normal weighting <i>Gerrit Grobler</i>	Analysis of interdependence between agricultural and energy commodity price dynamics with Bayesian multivariate DCC GARCH approach <i>Yegnanew A Shiferaw</i>		Unified approach in evaluating Shewhart_Xbar control charts with w-of-w supplementary runs-rules <i>Sandile Shongwe</i>	Evaluation of Selection Tools for the Inefficiency Distribution in Stochastic Frontier Models <i>Aviwe Gqwaka</i>

14h50 – 15h10	Pearson Demonstration	Occupancy models using logit link functions <i>Allan Clark</i>	Analysis Of Dropout Rates For Masters Students At The University Of Limpopo <i>Modupi Peter Mphekgwana</i>	Tolerance intervals for photovoltaic energy output assessment <i>Jani Deyzel</i>	Visualising Multidimensional Data using a Principal Surface Biplot <i>Raeesa Ganey</i>
15h10 – 15h40	Tea and Coffee				
15h15	Bayes Group Meeting (Venue: NB Pityana Room 02-56)				
15h40 – 16h40	Plenary talk Venue: Thamsanqa Kambule (Auditorium) Chair: Dr Freedom Gumedze Characterizing growth patterns based on longitudinal studies in children <i>Prof Louise Ryan</i>				
16h40 – 17:30	SASA Annual General Meeting Venue: Thamsanqa Kambule Auditorium				
18:00 – 21:00	Young Statisticians Braai				

Wednesday, 28th November 2018

	Morning tea and coffee				
	Stream 11: Special Session in Honour of Prof Crowther Venue: Council Chambers Chair: <i>Prof Legesse Debusho</i>	Stream 12: Biostatistics Venue: Thamsanqa Chair: <i>Dr Humphrey Brydon</i>	Stream 13: Development of Spatial Statistics Venue: Room 2-57 Chair: <i>Dr Inger Fabris-Rotelli</i>	Stream 14: General and Applied Statistics Venue: Room 301 Chair: <i>Prof John Olaomi</i>	Stream 15: Young Statisticians Venue: Bram Fischer Chair: <i>Mr Mark de Lancey</i>
09h00 – 09h20	Avifaunal Statistics at Wind Farms <i>Nico Laubscher</i>	Estimating Causal Effects in Health Outcomes Observational Studies: A Rank-Based Mahalanobis Distance Weighting Approach <i>Lateef Amusa</i>	A New Method To Analyse Multivariate Spatial Autocorrelation <i>Timotheus Darikwa</i>	Segmentation for estimating household rooftop solar installations <i>Jenny Holloway</i>	Joint modeling of recurrent events and survival outcomes using a Bayesian non-parametric Dirichlet Process Mixture priors approach <i>Adeboye Azeez</i>

09h25 – 09h45	The Biochemistry Of Chronic Fatigue: A Statistical Study <i>Francois Steffens</i>	Factors Associated with Mortality Among Under-five Children: Illustration Using Rwanda Demographic Health Survey Data, 2015 <i>Welcome Jabulani Dlamini</i>	Spatial Analysis Of Winter Rainfall Variability In Western Cape, South Africa <i>Willard Zvarevashe</i>	General Solutions To Robust Estimation Of Autoregressive Parameters <i>Simon Setsweke Nkoane</i>	Species Distribution Modelling: A Case Study for Tsetse Flies <i>Nada Abdelatif</i>
09h50 – 10h10	MLEC for parameter structures in the exponential class <i>Nina Strydom</i>	An Adaptive Simplified Acute Physiology Score 3 For The Prediction Of In-Hospital Mortality <i>Sisa Pazi</i>	Inhomogenous spatial modelling of DPT pulses for marine images <i>Inger Fabris-Rotelli</i>	Patterns of aggregation in count data with application to helminth egg counts <i>Phuti Sebatjane</i>	Evidence-based extraction and uncertainty evaluation of informal roads from remote sensing images <i>Renate Thiede</i>
10h15 – 10h35	The analysis of grouped data <i>Gretel Crafford</i>	Identification of features associated with sexual activity amongst first-year incoming students <i>Humphrey Brydon</i>		Modelling volatility of the JSE stock market with the EGARCH and Beta-Skew-t-EGARCH <i>Lesego Sepato</i>	Usage of the CLARA clustering algorithm in conjunction with the Discrete Pulse Transform for fast detection of block pulses in big data <i>Mark de Lancey</i>
10h40 – 11h15	Tea and Coffee				
	Stream 16: Regression Quantiles Venue: Room 301 Chair: <i>Mr Innocent Mudhombo</i>	Stream 17: Biostatistics Venue: Thamsanqa Chair: <i>Ms Esme Jordaan</i>	Stream 18: Multivariate Data Analysis Venue: Room 2-57 Chair: <i>Prof Ian MacDonald</i>	Stream 19: Categorical Data Analysis Venue: Council Chambers Chair: <i>Prof Kamanzi Binyanvanga</i>	Stream 20: Young Statisticians Venue: Bram Fischer Chair: <i>Ms Sonelca Muller</i>
11h15 – 11h35	Quantile interval regression models <i>Legesse Kassa Debusho</i>	Do not dictate a Survival Model to your Data with application to Malaria time to Re- infection Dataset <i>Ruffin Mpiana Mutambayi</i>	KNN-Triplot classification with polybags <i>Carel van der Merwe</i>	Application of multivariate generalised autoregressive score model on JSE and SSE all share index: A Copula approach <i>Katleho Makatjane</i>	Modelling Payment Dates Using A Mixture Of Bernoulli Distributions Approach <i>Lerato Langa</i>
11h40 – 12h00	Extending Statistics of the Hat Matrix to Regression Quantiles via the Elemental Regression Method <i>Edmore Ranganai</i>	Spatial and Temporal Distribution of HIV and TB <i>Annah Managa</i>	A Review and Application of Multivariate Causal Inference in Observational Data <i>Halima Twabi</i>	Joint modeling of correlated binary outcomes using structural equation modelling <i>Robert Keli</i>	Multiscale spatial modeling of texture images <i>Carel van Niekerk</i>
12h05 – 12h25	Pseudo-Likelihood and Estimating Equation	Smooth estimation of the baseline hazard function in a discrete survival	Comparison of Unfolding and Correspondence Analysis <i>Sugnet Lubbe</i>	Classification of Under-five Mortality: Comparison of Logistic Regression with Machine learning Methods	A Goodness-Of-Fit Test For The Rayleigh Distribution Based On A Lesser-Known Characterisation

	Methodology for Handling Missing Data <i>Birhanu Teshome Ayele</i>	model with application to age at first alcohol intake <i>Mahlageng Mashabela</i>		<i>Samuel Oduse</i>	<i>Shawn Liebenberg</i>
12h30 – 12h50	Multicollinearity Diagnosis in Regression Quantiles <i>Innocent Mudhombo</i>	Latent class, local independence and model misspecification <i>Esme Jordaan</i>	Maximum-likelihood estimation for multivariate exponential and Weibull distributions of Marshall-Olkin type <i>Iain MacDonald</i>	Summation over Partitions of Index Sets <i>Kamanzi Binyavanga</i>	Review Of Confidence Interval Estimators For Binomial <i>Sonelca Muller</i>
12h55 – 14h00	Lunch				
13:00 – 17:00	Field Excursions—Trip to the Cradle of Humankind and to Soweto Tour of Science campus				
13h30	IBS-Group South Africa Meeting (Venue: Bram Fischer)				
15h00	Data Science Group Meeting (Venue: Thamsanqa Kambule)				
19h00 –	Conference Dinner 19h00 for 19h30 until late				

Thursday, 29th November 2018

	Morning tea and coffee				
	Stream 21: Official Statistics and Public Policy Venue: Room 301 Chair: <i>Prof John Olaomi</i>	Stream 22: Biostatistics Venue: Thamsanqa Chair: <i>Dr Freedom Gumedze</i>	Stream 23: Multivariate Data Analysis Venue: Room 2-57 Chair: <i>Prof Peter Njuho</i>	Stream 24: Experimental Design/Ecology/Biometry Venue: Council Chambers Chair: <i>Prof Paul Fatti</i>	Stream 25: Stochastic Processes Venue: Bram Fischer Chair: <i>Dr Delson Chikobvu</i>
09h00 – 09h20	Is it a rip-off or not? The changes in the service delivery patterns for the rural nodes in South Africa <i>Malibongwe Christopher Mhemhe</i>	Geostatistical modelling of childhood anaemia in four sub-Saharan African countries <i>Danielle Roberts</i>	Improved Structural Equation Models Using Principal Component Analysis <i>Busanga Jerome Kanyama</i>	A semi-analytical approach for constructing D-optimal designs for the binary logistic model <i>Gaëtan Kabera</i>	Likelihood Inference and Jump Detection for Non-Linear Jump Diffusions With State-Dependent Intensity <i>Etienne Pienaar</i>
09h25 – 09h45	A second-order corrected confidence interval for the	Resampled Cox Proportional Hazards Models For Infant Mortality	Modelling BVI and BMI for body shape and size classification for men in South Africa	How to determine noise factors for calibration using the parallel fitting method	Modeling Sequential Adverse Events of HIV/AIDS: Transition-Specific Parametric Multi-state Model

	difference between two small area means <i>Yegnanew A Shiferaw</i>	At Kigali University Teaching Hospital <i>Paul Gatabazi</i>	<i>Busisiwe Tabo</i>	<i>Principal Ndlovu</i>	<i>Zelalem Dessie</i>
09h50 – 10h10	Modelling nitrogen dioxide (NO ₂) emission from Eskom's coal fired power station using generalised linear models (GLM) <i>Mpendulo Wiseman Mamba</i>	Analysis of Outlying and Influential Districts in the Modelling of Child Survival in Malawi <i>Tsirizani Kaombe</i>	Is the Mahalanobis distance appropriate for testing multivariate goodness-of-fit? <i>Wallina Oosthuizen</i>	The Graeco-Latin square and Hyper Graeco-Latin square designs <i>Henri Moolman</i>	HIV Disease Progression Based on CD4 Cell Count for Patients on ART in Zimbabwe: A Multistate Markov Model <i>Zvifadzo Matsena Zingoni</i>
10h15- 10h35		Diagnostics for detecting influential observations in joint survival models <i>Isaac Singini</i>	An Omnibus Test for Heteroscedasticity Using Radial Stationarity and Data Depth <i>Thomas Farrar</i>	Experimental Design for Interpreting Neural Network models <i>Paul Fatti</i>	A Markov model to estimate mortality due to HIV/AIDS using CD4 cell <i>Delson Chikobvu</i>
10h40 – 11h15	Tea and Coffee				
	Stream 26: Big Data and Data Science Venue: Room 301 Chair: <i>Mr Stefan Britz</i>	Stream 27: Statistics in Education Venue: Thamsanqa Chair: <i>Dr Lizelle Fletcher</i>	Stream 28: Experimental Design/Ecology/Biometry Venue: Council Chambers Chair: <i>Prof Edmore Ranganai</i>	Stream 29: Young Statisticians Venue: <i>Dr Kabera Gaetan</i> Chair: Bram Fischer	
11h15 – 11h35	Extracting Decision trees from Possibly Incomplete Causes of Death Data in South Africa <i>Bhekisipho Twala</i>	Toward Understanding the Needs of Introductory Statistics Students: An Interpretivist Study <i>Wilma Coetzee</i>	Smoothing large time series with quarter circular wavelets <i>Farai Mlambo</i>	The Differential Effect of Maternal Dietary Patterns on Quantiles of Birthweight <i>Aweke A. Mitku</i>	
11h40 – 12h00	Big data, what is all the fuss about? <i>Wessel Rheeder</i>	Insights into the Final Grade 12 Mathematics Examination Marks <i>Nombuso Zondo</i>	Social Network Analysis of Disease Transmission in a Social Carnivore <i>Victoria Goodall</i>	Evaluation of data quality, descriptive statistics and missing data using optimal multiple imputation methods in an ARV pharmacovigilance data in South Africa <i>Maria Lekganyane</i>	
12h05 – 12h25	Matching a Bullet to a Gun: One-Shot Learning and Convolutional Siamese Neural Networks <i>Stefan Britz</i>	Smartphone application for developing statistical literacy in South Africa's official languages <i>Tim Low</i>			
12h30 – 13h30	Plenary talk Venue: Thamsanqa Kambule (Auditorium)				

	Chair: Prof S Lubbe Classification and survival trees and forests <i>Prof Berthold Lausen</i>
13h30	Closing Ceremony Venue: Thamsanqa Kambule (Auditorium)
	Packed lunch and departure

Guidelines to poster presenters

There will be three different poster sessions, each from the morning until lunch time. Please check the programme later in this booklet (Section 11) to see on which day your poster session will be (they are arranged in the order the abstracts were submitted). Please put up your posters at the beginning of the day and remove them at the end of lunch time. The poster venue will be at the Exam hall (which is also the venue for exhibitions and tea and coffee breaks).

The poster boards are designed to take an A0 (841 x 1189 mm) size poster, in portrait orientation. Everything you need to fasten your posters will be made available, please contact the staff at the Registration desk if you need assistance.

Do feel free to contact the organisers if you have any queries at all!

Poster presentation sessions

Tuesday 27 November 2018

Helene Nel	Multilevel models for longitudinal data applied to an autism study
Thomas Farrar	When is the ideal time to play the South African National Lottery's LOTTO game?
Christine van Zyl	Capacity approximations for different fading distributions
Paige Els	Scaled average bioequivalence for classical four-period, two-sequence crossover design clinical trials
Gabriel de Wet	Learning conditional dependency structures within academic programmes using Bayesian Networks
Salomi Millard	Another look at self-consistency
Hartley Phasha	Types of texts that are best suited to the one topic per document assumption
Emile Fourie	Goodness-of-fit tests for normality
Kabelo Mahloromela	Selecting nonconvex windows for Spatial data
Rene Kirsten	Similarity of spatial point patterns
Neill Smit	Accelerated life testing for the generalised Eyring-Weibull model
Jean-Pierre Stander	Why stochastic gradient descent
Claire Bezuidenhout	GARCH models for time series exhibiting volatility
Amina Hoosain Mia	Statistics behind consumerism
Phumla Tsematse	Linear Versus Non-linear Versions of the Capital Asset Pricing Model
Tanya Potgieter	Testing and model specification of multilevel modeling
Bradley Paulse	Statistical Analysis of the Logistic regression using Complex Sampling Data

Wednesday 28 November

Frissiano Honwana	Applications of Linear Mixed Effects Models for Modelling Stem Radial Growth of Eucalyptus Trees
Thobeka Nombebe	Investigating the effect of restricting residuals in bootstrap based hypothesis testing
Selokela Victoria Molautsi	Modelling the sporadic behaviour of rainfall in the Limpopo Province, South Africa: an application of ETS state space and SARIMA models
Thabo Maifala	Statistical analysis of repeated measures data
Dzulani Mamphiswana	Mixture of generalised gamma distribution: Estimation
Itumeleng Kgobane	Image segmentation using a spatial variant of Gaussian mixture models.
Christopher Dunderdale	Image Classification in Photovoltaic Systems using Machine Learning Techniques
Sharday Nigrini	Bayesian Hierarchical Clustering of microarray gene expression data analysis
Juan-Pierre van der Merwe	Spatially weighted principal component analysis for identifying probabilistic seasonal prediction indicators.
Cairstine De Kock	Comparison of machine learning algorithms for predictive modelling of at-risk students
Daniel Mashishi	Modelling monthly average rainfall using r-largest order statistic for South Africa
Aleksandra Dabic	Product of independent kappa-mu type random variables with application in communication systems
Colette Le Roux	Gaussian Process Conditional Copula
Hannaline Roux	Wrapped generalised skew-normal type model with application
Makwelantle Asnath	Modelling malaria incidence in the Limpopo Province, South Africa: Comparison of classical and Bayesian methods of estimation.
Vongani Maluleke	Poverty Estimation Combining Aerial Images with Survey Data: An Application from South Africa

Thursday 29 November

Trystan Wragg	Bivariate Polya-Aeppli type II distribution
Henri Rothmann	K-function for big data
Lizalise Mngcele	Spatial Patterns Towards the Characterization of Karoo Environmental Features
limbasazethu Ngalo	Spatial analysis of factors associated with HIV infection among adults in Mozambique, 2015
Mapitsi Rangata	Exploring applicability of Functional Data Analysis methods to analyse space and time variations of climate data for South Africa
Carl Steyn	Addressing uncertainty: A Bayesian approach to deep learning
Tarisai Chimbwa	Longitudinal Multivariate Meta-Analysis Models
Mario Armindo Pumule	Spatial Representation Of South Africa Rainfall Variability
Vitumbiko Chijere Chirwa	Bayesian Spatial Modelling of Malaria risk in Malawian under-five children: evidence from 2017 Malawi Indicator Survey
Yusentha Balakrishna	Indoor Temperatures in Low Cost Housing in Johannesburg, South Africa
Wanda Ndamse	A Statistical Learning Approach to Customer Churn Prediction: a micro-insurance case
Shibe Mhlongo	Statistical methods to facilitate the comparison of the mental health status of rape exposed and non-exposed women one year post enrollment: Rape Impact Cohort Evaluation (RICE) study
Kgabo Kubyana	Circular models and their statistical inference
Thembinkosi Kunene	The cross-K function for marked spatial point patterns.
Andriase Ramochele	Investigating causal relationships of sunspot activity linked to streamflow in the Vaal Dam catchment area over time

Changes in the SASA 2018 Programme

Date	Time	Stream	New Talks	Cancelled Talks
Tuesday, 27 th	11h35-11h55	1	Big Data Credit Scoring <i>Dumisani Nkambule</i>	Probability models for predicting electrical system faults and estimating voltage quality loss by machine learning <i>Mphiliseni Nzuza</i>
Wednesday, 28 th	09h25-09h45	14		General Solutions to Robust Estimation of Autoregressive Parameters (Move to Poster) <i>Simon Setsweke Nkoane</i>
Wednesday, 28 th	11h15-11h35	20	Ensemble methods in multi-label classification <i>Annegret Muller</i>	Modelling Payment Dates Using A Mixture of Bernoulli Distributions Approach (Move to Poster) <i>Lerato Langa</i>
Thursday, 29 th	12h05-12h25	28	Survival Analysis and Business Applications: The case Studies <i>Ewa Fraczak</i>	

Abstracts for the New Talks

	Big Data Credit Scoring <i>Dumisani Nkambule</i>
Tuesday, 27 th (11h35-11h55) Stream 1	<p>Credit scoring is a form of artificial intelligence based on predictive modeling that assesses the likelihood of a customer defaulting on a credit obligation, becoming delinquent or insolvent. The greatest benefit of credit scoring is the ability to help make decisions in a fast and efficient way, such as to accept or reject a customer or increase or decrease loan value, interest rate, or term. This paper identifies and investigates one of many challenges that occur during the development of credit scorecards. We examine the performance of a number of state-of-the-art classification techniques on an imbalanced credit scoring Big dataset. We further show how artificial data may be used to overcome difficulties associated with class imbalances in Big Data.</p>
	Ensemble methods in multi-label classification <i>Annegret Muller</i>
Wednesday, 28 th (11h15-11h35) Stream 20	<p>There are many scenarios where several labels may be associated simultaneously with each data case in a dataset. Therefore, a large number of multi-label datasets are found in a variety of domains, including image annotation, text annotation, bioacoustics, music research and medical diagnostics. With such datasets becoming increasingly available, multi-label classification has become an active area of research. Multi-label classification is an extension of binary- and multi-class classification to scenarios where each instance in a dataset can have multiple or none of K labels.</p> <p>Methods to perform multi-label classification can be divided into three categories: problem transformation methods, algorithm adaptation methods and multi-label ensemble methods. Multi-label ensemble methods receive particular attention in this research. Previously proposed multi-label ensemble methods are described, namely ensembles of classifier chains, random forests of predictive clustering trees and random k-labelsets. A new multi-label ensemble method, named label dependent splitting (LDsplit) with trees, is also proposed.</p> <p>LDsplit with trees constructs an ensemble of tree-structures by considering different permutations of the labels in a multi-label dataset. The method differs from other multi-label ensemble methods since each tree-structure splits the multi-label data in a label-dependent way, whilst incorporating label correlation. Furthermore, each split of a node in a tree-structure is performed by considering a binary classification problem. By performing an empirical study on benchmark datasets, the predictive performance of LDsplit with trees is compared to that of other multi-label learning methods. LDsplit with trees produces very promising results, allowing us to believe that with further modifications the procedure may become a competitive multi-label learning method.</p>
	Survival Analysis and Business Applications: The case Studies <i>Ewa Fraczk</i>
Thursday, 29 th 12h05-12h25 Stream 28	<p>In the era of predictive models for business analyzes, the usability of modeling based on the use of survival analysis is growing. The SAS Book <i>Business Survival Analysis using SAS</i> (J. Ribeiro 2017) proves: the development and applications of survival techniques in business areas, such as: finance, telecommunication, insurance and many other areas of the industry is growing. The text of our paper contains examples of three business modeling applications based on survival analysis:</p> <p>First - modeling the bankruptcy of enterprises using non-parametric models.</p> <p>The second - default modeling using semiparametric models</p> <p>The third - modeling clients migration process between segments in a service company using a survival recurrent model.</p>