Onderwerpen Masterproef Opleiding Wiskunde
Academiejaar 2023-2024
Vakgroep Toegepaste Wiskunde, Informatica en Statistiek

Waardebepaling van financiële en verzekeringsproducten via Fouriertransformatiemethoden
Promotor: Prof. dr. M. Vanmaele
Doelgroep: studenten uit de tweede master Wiskunde

Korte beschrijving
Efficiënte numerieke methoden zijn belangrijk om snel de prijs van complexe (financiële en verzekerings-)contracten en verschillende financiële modellen te kalibreren. De bestaande numerieke methoden kunnen in drie grote groepen ingedeeld worden: methoden voor partiële (integraal-)differentiaalvergelijkingen (PIDV), Monte Carlo-simulatie en numerieke integratie-methoden.
Wanneer de Fouriertransformatie van de dichtheidsfunctie beschikbaar is, vormen Fourier-gebaseerde methoden een zeer competitief alternatief voor andere numerieke technieken. Het onderwerp van deze masterproef is de studie van recente numerieke en theoretische ontwikkelingen van Fouriertransformatiemethoden om verschillende uitdagingen in de financiële en verzekeringssector aan te gaan. De uitdagingen en toepassingen kunnen variëren van de efficiënte prijsbepaling van multi-asset opties, de prijsbepaling van opties onder uitdagende nieuwe modellen, de waardering van complexe levensverzekeringscontracten en de efficiënte berekening van risicomaatstaven en risicoblootstellingen voor portefeuilles.

Mogelijke artikels:

- Laura Ballotta, Ernst Eberlein, Thorsten Schmidt, Raghid Zeineddine (2021) Fourier based methods for the management of complex life insurance products, Insurance: Mathematics and Economics, Volume 101, Part B, 320-341
  https://doi.org/10.1016/j.insmatheco.2021.08.009
- Fang Fang, Xiaoyu Shen and Chujun Qiu (2022) A New and Efficient Fourier Method for Risk Quantification and Allocation of Credit Portfolios. Available at SSRN: http://dx.doi.org/10.2139/ssrn.4153482
- Yayun Wang, Zhimin Zhang, Wenguang Yu (2021) Pricing equity-linked death benefits by complex Fourier series expansion in a regime-switching jump diffusion model, Applied Mathematics and Computation, Volume 399, 126031
  https://doi.org/10.1016/j.amc.2021.126031
- Weinan Zhang, Pingping Zeng, Gongqiu Zhang (2023) Pricing Discretely Monitored Asian Options under Regime-Switching and Stochastic Volatility Models with Jumps: Fourier-Cosine Approach, preprint
- Wei Zhong, Zhenyu Cui, Zhimin Zhang (2023) Efficient valuation of guaranteed minimum maturity benefits in regime switching jump diffusion models with surrender risk, Journal of Computational and Applied Mathematics, Volume 422, 114914 https://doi.org/10.1016/j.cam.2022.114914

Hedgen van basisrisico met behulp van reinforcement learning
Promotor: Prof. dr. M. Vanmaele
Doelgroep: studenten uit de tweede master Wiskunde

Korte beschrijving
Basisrisico wordt gedefinieerd als het inherente risico dat een handelaar neemt bij het afdekken van een positie door een tegengestelde positie te nemen in een derivaat van het actief, zoals een futurescontract. Dit basisrisico wordt aanvaard in een poging het prijsrisico af te dekken. Bijvoorbeeld, als de huidige spotprijs, of contante marktprijs, van goud 1190 EUR is, en de prijs van goud in het goudtermijncontract in juni 1195 EUR bedraagt, is de basis, het verschil, 5,00 EUR. Het basisrisico is het risico dat de prijs van de futures niet beweegt in normale, gestage correlatie met de prijs van

de onderliggende waarde en dat deze fluctuatie in de basis de effectiviteit van een afdekkingsstrategie (die wordt gebruikt om de blootstelling van een handelaar aan mogelijk verlies te minimaliseren) teniet kan doen. Het verschil tussen de contante prijs en de prijs van de futures kan ofwel groter of kleiner worden tussen het moment waarop een afdekkingspositie wordt ingenomen en het moment waarop die positie wordt gesloten.
Er zijn verschillende types basisrisico te onderscheiden namelijk basisrisico verbonden aan de prijs, de locatie, de kalender of de kwaliteit van het product.
De bedoeling van deze masterscriptie is om hedgingstrategieën te bestuderen in de aanwezigheid van basisrisico en deze te vergelijken met een hedgingstrategie gebaseerd op reinforce-ment learning. Reinforcement learning (RL) RL is een gebied van machine learning dat zich bezighoudt met hoe intelligente actoren acties moeten ondernemen in een omgeving om een cumulatieve beloning te maximaliseren. Hiervoor kan er voortgebouwd worden op een masterproef van 2020-2021 waar een RL algoritme op basis van "deep deterministic policy gradient" is gebruikt.

Relevante literatuur vormen de volgende artikels:

- Monoyios, M. (2004). Performance of utility-based strategies for hedging basis risk. Quantitative Finance, 4(3), 245–255.
- Trottier, D. A., Godin, F., & Hamel, E. (2018). Local hedging of variable annuities in the presence of basis risk. ASTIN Bulletin, 48(2), 611–646.
- Axel F.A. Adam, Ingmar Nolte (2011). Cross hedging under multiplicative basis risk. Journal of Banking & Finance, 35:2956–2964.
- Broll, U., Welzel, P. & Wong, K. P. (2015) Futures hedging with basis risk and expectation dependence. International Review of Economics, 62(3):213–221.
- Sutton, R. & Barto, A.G. (2018) Reinforcement learning: An introduction, MIT press
- Xue, X., Zhang, J. & Weng, C. (2019) Mean-variance hedging with basis risk. Applied Stochastic Models in Business and Industry, 35:704–716

Stein's lemma en toepassingen in portefeuilleselectieproblemen

Promotor: Prof. dr. M. Vanmaele
Doelgroep: studenten uit de tweede master Wiskunde

Korte beschrijving

Voor een bivariate normaalverdeelde vector $(X , Y)^T$, heeft Stein (1973, 1981) aangetoond dat Cov [h (X) , Y] = Cov [X, Y] E [h ' (X)] voor een willekeurige afleidbare functie h waarvoor E[| h '(X)|] bestaat.
Dit lemma heeft zijn nut bewezen in heel wat disciplines, zoals statistiek, waarschijnlijkheid, besliskunde en financiën. Rendementen van aandelen zijn echter niet altijd symmetrisch verdeeld, ze vertonen scheefheid.
Deze observatie zette Adcock ertoe aan om in een serie van artikels een type van Stein's lemma op te stellen voor multivariate verdelingen die consistent zijn met Simaan's (1993) modelering van aandelenrendementen.
Landsmann en co-auteurs hebben verdere uitbreidingen van Stein's lemma bewezen voor (multivariate) elliptisch verdeelde veranderlijken met toepassingen in o.a. risicotheorie (zie Landsmann & Valdez, 2016).
Vanduffel en Yao (2017) vertrekken van Simaan's setting en stellen een nieuw type van Stein's lemma op voor de multivariate veralgemeende hyperbolische verdeling. Als toepassing, beschouwen ze een portefeuilleselectievraagstuk.
De opdracht van deze masterproef bestaat in het bestuderen van Stein's lemma voor verschillende verdelingen en de toepassingen ervan in portefeuilleselectieproblemen en bij risicomaten.
Voor deze studie kan gesteund worden op de volgende artikels:

- C.J. Adcock (2014). Mean–variance–skewness efficient surfaces, Stein's lemma and the multivariate extended skew-Student distribution. European Journal of Operational Research, 234:392–401
- Z. Landsmann and E. Valdez (2016). The tail Stein's identity with applications to risk measures. North American Actuarial Journal, 20(4): 313-326.
- Y. Simaan (1993). Portfolio selection and asset pricing-three parameter framework. Management Science, 39 (5):568–587.
- C.M. Stein (1973). Estimation of the mean of a multivariate normal distribution. In Proceedings of the Prague symposium on asymptotic statistics (pp. 345–381) .

- C. M. Stein (1981). Estimation of the mean of a multivariate normal distribution. Annals of Statistics, 9:1135–1151.
- S. Vanduffel and J. Yao (2017). A Stein type lemma for the multivariate generalized hyperbolic distribution. European Journal of Operational Research, 261(2):606-612.

Een ander onderwerp in het domein van de financiële of actuariële wiskunde
Promotor: Prof. dr. M. Vanmaele

Korte beschrijving:
Bespreekbaar
Doelgroep: studenten uit de tweede master Wiskunde

Titel: Factor analyse
Promotor: Prof. Stijn Vansteelandt

Korte beschrijving:
Factor analyse is een zeer populaire techniek om metingen te representeren die moeilijk te meten zijn, zoals intelligentie, levensvoldoening, … Hiertoe worden meerdere metingen opgevraagd die iets vertellen over intelligentie, levensvoldoening, … en die worden vervolgens op een slimme manier samengevat in een zogenaamde 'factor score'.
In deze masterproef zullen we factor analyse bestuderen en vervolgens ingaan op recent werk waar aangetoond wordt dat dergelijke factor analyses problematisch kunnen zijn wanneer de factor score in verdere analyses verwerkt wordt. Hiertoe werden statistische testen ontwikkeld om na te gaan of het model achterliggend aan een factor analyse de data tegenspreekt.
We zullen de performantie van deze testen onderzoeken in simulatiestudies en ze vervolgens toepassen op gegevens rond levensvoldoening bekomen voor bijna 13000 deelnemers aan de Health and Retirement Study.
Afhankelijk van de voorkeur van de student kan een doelstelling er verder in bestaan om de voorgestelde testen uit te breiden om ze meer flexibel te maken.

Titel: Statistische besluitvorming na gebruik van machine learning technieken
Promotor: Prof. Stijn Vansteelandt

Korte beschrijving:
Meer en meer statistische analyses maken gebruik van machine learning technieken om bijvoorbeeld het effect van een behandeling of interventie te schatten. Zoals alle statistische analyses, genereren ook machine learning technieken onzekerheid op de resultaten (ten gevolge van steekproefvariatie);
omwille van de complexiteit van de technieken wordt dit in de verdere analyse echter meestal genegeerd. Als gevolg hiervan zijn de betrouwbaarheidsintervallen die men bekomt, typisch zwaar vertekend.
In deze masterproef zult u betrokken worden in recente ontwikkelingen, alsook onderzoek binnen de onderzoeksgroep van de promotor, om na te gaan hoe men hiermee kan omgaan. De bestudeerde methodes zullen theoretische en/of door middel van simulatiestudies en concrete data-analyses worden geëvalueerd.

Titel: Een ander onderwerp in het domein van de mathematische statistiek en/of toegepaste data-analyse
Promotor: Prof. Stijn Vansteelandt

Korte beschrijving:
Bespreekbaar

Titel: Algoritmen voor het Steiner Tree Problem
Promotor: Prof. Veerle Fack

Korte beschrijving:
In een brede betekenis bestaat het doel van een Steiner Tree Problem erin om de goedkoopste manier te bepalen om een set van objecten te verbinden.
In de meest voorkomende varianten zijn deze objecten ofwel punten in een metrische ruimte ofwel een deelverzameling van de toppen van een netwerk/graaf, en het doel is het bepalen van een boom die ze allemaal verbindt.
Er zijn talloze toepassingen van deze problemen, zoals netwerkoptimalisatie, reconstructie van phylogenetische bomen, ontwerp van computercircuits, multicast-routering in communicatienetwerken.
Dit probleem was ook het onderwerp van een DIMACS Implementation Challenge in 2014, waarbij o.m.
een collectie benchmark data opgesteld werd, naast het uitwerken van gevorderde algoritmen voor het aanpakken van dit probleem.
Met deze Implementation Challenge als startpunt, is het de bedoeling van deze masterproef om bestaande algoritmen voor het probleem te bestuderen en te analyseren, evenals eigen benaderingen uit te werken en te toetsen tegenover de bestaande methodes

---

Titel: Fractionele Sturm-Liouville vergelijkingen
Promotor: Prof. Marnix Van Daele

Korte beschrijving:
In de cursus Wiskundige modellering hebben we het in het laatste deel gehad over het (klassieke) Sturm-Liouvilleprobleem (SLP).
Dit is een eigenwaardeprobleem waarbij de oplossing y(x) van een tweede-orde differentiaalvergelijking (p(x) y')' + q(x) y = E w(x) y ook dient te voldoen aan bepaalde randvoorwaarden.
Heel veel problemen, uit zowat alle takken van de wetenschappen, kunnen uitgedrukt worden als een klassiek SLP. Maar in de laatste decennia is ook gebleken dat veel wetenschappelijke problemen nog accurater kunnen gemodelleerd worden door gebruik te maken van fractionele calculus i.p.v. de klassieke calculus, waarbij slechts afgeleiden van gehele orde bestaan. Zo ontstaan Fractionele Sturm-Liouvilleproblemen.

Het doel van dit thesisonderwerp is tweevoudig:
- deel 1: kennis te maken met fractionele calculus: hoe wordt bvb. de afgeleide van orde r gedefinieerd als r geen natuurlijk getal is?
- deel 2: fractionele calculus toe te passen op een specifiek probleem zoals het fractionele Sturm-Liouville probleem.

Start punt:
- deel 1: een boek rond fractionele calculus, bvb: Igor Podlubny: Fractional differential equations. Mathematics in Science and Engineering Vol. 198 (London, Academic Press, 1999)
- deel 2: artikels rond Fractionele Sturm-Liouville vergelijkingen, zoals bvb. Dehghan, M.; Mingarelli, A.B. Fractional Sturm-Liouville eigenvalue problems, I. Rev. Real Acad. Cienc. Exactas Fis. Nat. Ser. A Mat. 2020, 114, 46.

Titel: Ontwikkeling van een kwaliteitscontrole systeem voor kwantum gegenereerde randomgetallen.
Promotor: Prof. dr. Marnix Van Daele;
Begeleider: Filip Pynckels (Auditeur-generaal ICT, FOD Binnenlandse Zaken)

Korte beschrijving:
Op basis van de studie Cryptosystems in an ever-changing world kan besloten worden dat een goede encryptie steunt op een goede hardware randomgenerator die niet te beïnvloeden is door externe invloeden. Vaak vindt men dergelijke

hardware ingebed in de CPU van een computersysteem, maar deze kan door bijvoorbeeld elektromagnetische "*fout injectie*" bijgestuurd worden in de richting van voorspelbaarheid. Een "*proof of concept*" randomgetallen generator op basis van "*kwantum tunnel effect*" werd ontwikkeld (zie onderstaande foto). Dit is evenwel nog maar de eerste stap in een volledig ontwikkeltraject (hardware aanpassingen, verkleinen, softwareontwikkeling en optimalisatie, kwaliteitscontrole van de gegenereerde randomgetallen, EMP-resistent maken, design van een cover, schrijven van documentatie, …).



De volgende stappen zijn:

- Een correcte methode ontwikkelen (theoretisch en programmatie van een controller met slechts 8Kb geheugen waarvan 2Kb voor variabelen) om de gegenereerde entropie om te zetten in randomgetallen die via de USB-poort naar een computersysteem worden gestuurd.
- Device drivers ontwikkelen (Windows, macOS en Linux) die gebruik maken van een correcte te ontwikkelen methode (theoretisch en programmatie) om de gegenereerde randomgetallen te ontvangen, snel fulltime te controleren op kwaliteit en door te geven aan andere programma's op het computersysteem.

Bovenstaande stappen kunnen echter maar worden voltooid zodra er een standalone kwaliteitscontrole systeem voor randomgetallen beschikbaar is dat toelaat om de uitvoer van de boven beschreven stappen te controleren. Tijdens de ontwikkelfase van het project kan dit controle systeem dienen om de ontwikkeling te toetsen op correctheid, tijdens het gebruik van het eindproduct zal het systeem toelaten aan de gebruiker om eender wanneer na te gaan dat de randomgetallen generator correct werkt.

Voor deze scriptie wordt van de student verwacht dat hij/zij:

- Bestaande kwaliteitscontrole systemen en/of methodes voor gegenereerde randomgetallen vindt en bestudeert.
- Een goed kwaliteitscontrole systeem voor gegenereerde randomgetallen beschrijft en ontwikkelt of verder uitbouwt.

Meer specifiek wordt van de student verwacht dat hij/zij:

- Kennis opdoet van kwantum entropie generatie, elektromagnetische interferentie, elektromagnetische fout injectie, methodes om het hardware systeem te compromitteren, … (Opleiding van maximaal 2 uur).
- Kennis opdoet van de hardware aspecten van de hardware entropie generator in zijn huidige vorm (Opleiding van maximaal 1 uur).
- Bestudeert aan welke vereisten kwalitatieve random sequenties moeten voldoen (zelfstudie met overlegmomenten).
- Bestudeert welke de sterktes en zwaktes van bestaande kwaliteitssystemen zijn (zelfstandig werk met overlegmomenten).
- Een bestaand kwaliteitssysteem kiest en aanpast of er één ontwikkelt (zelfstandig werk met overlegmomenten).
- Het systeem dat hij/zij gekozen of gemaakt heeft verdedigt op basis van de zwaktes/sterktes ten opzichte van niet gekozen systemen (zelfstandig werk met overlegmomenten).
- Een korte beschrijving maakt hoe het door de student gekozen of ontwikkelde systeem moet gebruikt worden (zelfstandig werk met overlegmomenten).
- Het bovenstaande verwerkt in een samenvattende scriptie.

Referentie:

- "Cryptosystems in an ever-changing world", Freya Verbeke (begeleid door Prof. Marnix Van Daele, UGent, Faculteit Wetenschappen, Vakgroep Toegepaste Wiskunde, Informatica en Statistiek), 8 augustus 2019
- "Random number generator ver 2.0", Robin Pynckels & Filip Pynckels, 20 februari 2021 (niet voltooid document), https://pynckels.github.io/2020-4-Random_number_generator_ver_2-0.pdf

Titel: Een ander onderwerp in het domein van numerieke wiskunde
Promotor: Prof. Marnix Van Daele

Korte beschrijving:
Bespreekbaar
Doelgroep: Studenten uit de tweede master Wiskunde

Titel: Afhankelijke kansvariabelen en hun som
Promotor: Prof. David Vyncke

Korte beschrijving:
Het risico waaraan een financiële of verzekeringsportefeuille blootgesteld is, hangt nauw samen met de verdeling van een som van kansvariabelen. Indien de kansvariabelen onafhankelijk zijn, kan men die som schrijven als een convolutie, maar een dergelijke voorwaarde is in de praktijk zelden voldaan. Rekening houden met de reële afhankelijkheidsstructuur brengt echter heel wat problemen met zich mee. Arbenz et al (2011) ontwierpen een algoritme dat de verdeling snel zou moeten berekenen, maar dat algoritme is beperkt tot positieve kansvariabelen en vereist bovendien dat de volledige copula gekend is. In de praktijk is echter vaak slechts gedeeltelijke informatie over de afhankelijkheid bekend. Door gebruik te maken van dergelijke informatie is het mogelijk om de (verdeling van de) som te begrenzen, zoals geïllustreerd in Bernard et al (2017) en Lux & Papapantoleon (2019). In deze masterproef bestudeert de student verscheidene technieken om de verdeling van een som van afhankelijke kansvariabelen te berekenen en/of te begrenzen.

Referentie:

- Arbenz P., Embrechts P. & Puccetti G. (2011). The AEP algorithm for the fast computation of the distribution of the sum of dependent random variables. Bernoulli 17(2), 562–591.
- Bernard C., Rüschendorf L. & Vanduffel S. (2017). Value-at-risk bounds with variance constraints. Journal of Risk and Insurance 84, 923–959.
- Lux T. & Papapantoleon A. (2019). Model-free bounds on Value-at-Risk using extreme value information and statistical distances. Insurance: Mathematics and Economics 86, 73-83.

Titel: Een ander onderwerp in de financiële of actuariële wiskunde
Promotor: Prof. David Vyncke

Korte beschrijving:
Bespreekbaar
Doelgroep: Studenten uit de master Wiskunde

Titel: Similarity learning for fuzzy rough rule induction
Promotor: Prof. Chris Cornelis

Korte beschrijving:
Fuzzy Rough Sets (FRS) are a construct from artificial intelligence that has been successfully applied in various machine learning tasks, including feature selection, instance selection, classification, and regression [1].

FRS make use of a similarity relation, which expresses the degree to which two elements in a dataset are related. Instead of using a normal, fixed similarity relation, it is possible to apply similarity learning or distance metric learning (DML)[2] to learn the optimal relation from the data.
This has been successfully applied to neighborhoud-based machine learning methods, such as k-nearest neighbours.
Rule induction is a machine learning model which learns rules from the data with the aim of e.g. predicting the class of new samples. One such technique which makes use of FRS theory, is QuickRules [3], which creates rules in a greedy way.
The goal of this thesis is to design and conduct an extensive experimental study on a collection of benchmark datasets, in order to evaluate the predictive and descriptive potential of similarity learning in the context of rule induction.
The specific research questions that the thesis should address are:
• How can we apply DML to rule induction?
• Which DML methods perform best in this context, what are their strengths and weaknesses w.r.t. predictive and descriptive performance?
• Can we improve these methods or create our own technique which outperforms them?
• Are there specific kinds of data (e.g., imbalanced data, high-dimensional data, multi-label data, numerical vs categorical data, …) on which the approach performs better or worse?
Prior knowledge of fuzzy rough sets is not necessary, and a Python implementations of QuickRules will be provided. On the other hand, experience with setting up machine learning experiments is highly recommended.

Referentie
[1] S. Vluymans, et al. "Applications of Fuzzy Rough Set Theory in Machine Learning: a Survey." Fundamentae Informaticae 142.1-4 (2015): 53-86.
[2] J. L. Suárez, S. García, F. Herrera, A tutorial on distance metric learning: Mathematical foundations, algorithms, experimental analysis, prospects and challenges, Neurocomputing 425 (2021) 300–322.
[3] Richard Jensen, Chris Cornelis, and Qiang Shen. Hybrid fuzzy-rough rule induction and feature selection. pages 1151 – 1156, 09 2009.

Titel: Experimental evaluation of new fuzzy rough set models for machine learning tasks
Promotor: Prof. Chris Cornelis

Korte beschrijving:
Fuzzy Rough Sets (FRS) are a construct from artificial intelligence that has been successfully applied in various machine learning tasks, including feature selection, instance selection, classification, and regression. [1]
A prominent example is fuzzy rough nearest neighbour classification (FRNN), which is a lazy learner that associates an upper and a lower approximation with each decision class and classifies test instances according to their membership in these.
This has a transparent interpretation: upper approximation membership encodes to what extent a test instance is similar to the training instances of a class, and so possibly belongs to this class, whereas lower approximation membership encodes to what extent a test instance is not similar to the training instances of other classes and so necessarily belongs to this class.

One of the most promising FRS models uses Ordered Weighted Averaging (OWA) operators in the calculation of the upper and lower approximations. Recently, a new class of fuzzy rough set models have been developed at Ghent University, based on fuzzy quantification models. [2]
The goal of this thesis is to design and conduct an extensive experimental study on a collection of benchmark datasets, in order to evaluate the machine learning potential of the newly proposed FRS models.
The specific research questions that the thesis should address are:

- Are the new fuzzy rough set models based on fuzzy quantification an improvement on the existing OWA FRS model for feature selection?
- What are the strengths and weaknesses of each FRS model?

- Are there specific kinds of data (e.g., imbalanced data, high-dimensional data, multi-label data, numerical vs categorical data, …) on which the approach performs better or worse?

Prior knowledge of fuzzy rough sets is not necessary, and Python implementations of the different FRS models will be provided. On the other hand, experience with setting up machine learning experiments is highly recommended.

Referentie

[1] Vluymans, Sarah, et al. "Applications of Fuzzy Rough Set Theory in Machine Learning: a Survey." Fundam. Informaticae 142.1-4 (2015): 53-86.
[2] Theerens, Adnan, and Chris Cornelis. "Fuzzy Rough Sets Based on Fuzzy Quantification." arXiv preprint arXiv:2212.04327 (2022).

Titel: Normalised outlier scores and their application in robust fuzzy rough sets
Promotor: Prof. Chris Cornelis

Korte beschrijving:

Fuzzy Rough Sets (FRS) are a construct from artificial intelligence that has been successfully applied in various machine learning tasks, including feature selection, instance selection, classification, and regression [1]. However, outliers in the dataset can have a significant impact on the performance of the model. To overcome this issue, outlier detection techniques have been introduced, where outlier scores are assigned to each data point. Outlier scores are often normalised to improve the interpretability of results. Since these normalised outlier scores can be viewed as fuzzy sets, they are well fitted to be combined with fuzzy rough sets. However, the effectiveness of normalised outlier scores in robust fuzzy rough set models is yet to be explored. The objective of this thesis is to investigate the application of normalised outlier scores in robust fuzzy rough set models.
Specifically, the goal is to:

- Conduct a comprehensive review of the literature on normalised outlier scores.
- Develop a framework that integrates normalised outlier scores into robust fuzzy rough sets.
- Implement the proposed framework on benchmark datasets and evaluate its effectiveness in improving the accuracy of classification based on fuzzy rough sets.

Two potential frameworks for integrating normalised outlier scores are readily available: Choquet-based fuzzy rough sets [2] and confidence-based fuzzy quantifier FRS [3].
Prior knowledge of fuzzy rough sets is not necessary, and Python implementations of the different FRS models will be provided. On the other hand, experience with setting up machine learning experiments is highly recommended.

Referentie

[1] Vluymans, Sarah, et al. "Applications of Fuzzy Rough Set Theory in Machine Learning: a Survey." Fundam. Informaticae 142.1-4 (2015): 53-86.
[2] Theerens, Adnan, Oliver Urs Lenz, and Chris Cornelis. "Choquet-based fuzzy rough sets." Int. Journal of Approximate Reasoning 146 (2022): 62-78.
[3] Theerens, Adnan, and Chris Cornelis. "Fuzzy Quantifier-Based Fuzzy Rough Sets." 2022 17th Conference on Computer Science and Intelligence Systems (FedCSIS). IEEE, 2022.
[4] Theerens, Adnan, and Chris Cornelis. "Fuzzy Rough Sets Based on Fuzzy Quantification." arXiv preprint arXiv:2212.04327 (2022).

Titel: Statistical inference for optimal treatment regimes
Promotor: Prof. Oliver Dukes

Korte beschrijving:

Identifying the optimal treatment strategy or decision for a patient lies at the heart of personalized medicine. However, providing valid tests and confidence intervals in this setting turns out to be challenging when the optimal regime is non-unique. A simple way of obtaining valid inference is by sample-splitting; the optimal regime is learn on a training sample, and confidence intervals are conducted on a test sample. however, this wastes information, and results may depend heavily on the choice of split. The aim of this project is to construct confidence intervals for the mean outcome under the optimal strategy that improves upon naive sample-splitting.

Titel: Debiased machine learning with instrumental variables and binary outcomes
Promotor: Prof. Oliver Dukes

Korte beschrijving:
This project will consider inference for the causal effect of a binary exposure in a setting with a variable. This is a variable that has no direct effect on the outcome, is itself not subject to unmeasured confounding, and is predictive of the exposure. It can be used to infer the ATE even when the exposure-outcome relationship is distorted by unmeasured confounding. Existing results for statistical inference have relied on an additive homogeneity condition which is not usually plausible with a binary outcome. In this project, we will explore inference under a different assumption. We will begin by obtaining efficiency bounds under differing homogeneity conditions, and then understanding the asymptotic properties of corresponding 'debased machine learning' estimators. The considered methods will be evaluated theoretically and/or by means of simulation studies and concrete data analyses.