**(Insteek) Thierry Desot – ADS&AI – TC - overzicht weken**

**Datapunten met verwijzing naar geïntegreerde beroepspraktijk opdracht**

|  |  |  |
| --- | --- | --- |
| Week 1 | Reguliere expressies  (Datapunten: Classification/information extraction/chatbot) | Reguliere expressies om gericht informatie uit data te halen, data in een gewenst formaat voor traditionele machine learning en deep learning te gieten voor classificatie, interactieve dialoog, information extraction, data te cleanen en te preprocessen.  Kennis en toepassing van reguliere expressies zijn een vereiste voor alle datapunten en opdrachten voor beroepspraktijk.  **Voor de geïntegreerde opdracht moeten data (CVs) in een gewenst formaat gegoten, preprocessed en cleant worden d.m.v. reguliere expressies.**  Theorie, oefeningen, oefening in functie van de geïntegreerde opdracht |
| Week 2 | SpaCy  (Datapunten: Classification/information extraction/chatbot) | SpaCy voor sentence segmentation, tokenization, part-of-speech  Tagging, dependency parsing, NER  Python classes/objects  Voor de **geïntegreerde opdracht moeten voor gescreende CVs de belangrijkste kenmerken van een kandidaat profiel highlighted** zijn. Dat kan gebeuren d.m.v. SpaCy.  Theorie, oefeningen, oefening in functie van de geïntegreerde opdracht |
| Week 3 | Traditionele Machine learning  (Datapunten: Classification) | Emotie analyse met traditionele machine learning, logistic regression. Feature engineering met SpaCy. Train een ML model en evalueer accuracy, F score.  Pas deze traditionele machine learning techniek toe op Airline Sentiment analysis Twitter data set toe  Theorie, oefeningen, oefening in functie van de geïntegreerde opdracht  **Voor de geïntegreerde opdracht worden CVs geclassifieerd met een traditionele ML benadering.** |
| Week 4 | Traditionele Machine learning  (Datapunten: Classification) | Vergelijk performances uit vorige les, met gebruik van BOW, TF-IDF met SVM, decision trees, Naive Bayes.  **Voor de geïntegreerde opdracht worden CVs geclassifieerd met een traditionele ML benadering.** |
| Week 5 | Deep learning classifier  (Datapunten: Classification) | Deep learning classifier wordt getrained met 1 of meerdere data sets uit vorige lessen: RNN, LSTM  **Theorie, oefeningen, oefening in functie van de geïntegreerde opdracht, CV classificatie d.m.v. DL classifier** |
| Week 6 | Transformer modellen – pretrained – visualisatie  (Datapunten: Classification) | Transformer modellen worden toegelicht. Resultaten voor verscheidene data sets uit vorige lessen, gebruikt voor traditionele en DL ML benaderingen worden vergeleken bij gebruik van pre-trained transformer modellen, voor pre-trained sentiment analyse en zero-shot classification.  Linguïstische redeneerprocessen door het model worden gevisualiseerd in het kader van responsible technology/transparante AI.  Theorie, oefeningen, oefening in functie van de geïntegreerde opdracht  **De test sets voor de geïntegreerde opdracht worden hier toegepast.** |
| Week 7 | Transformer modellen – fine tuned  (Datapunten: Classification) | Een pre-trained transformer model wordt aangepast d.m.v. een custom data set. Hierbij worden meerdere data sets uit vorige lessen gebruikt en de resultaten vergeleken, voor de verschillende data sets (o.a. IMDB film data set).  In kader van responsible technology, wordt een tipje van de sluier van deep learning blackbox gelicht, door te analyseren welke word tokens vooral bijdragen aan correcte classificatie.  Theorie, oefeningen, oefening in functie van de geïntegreerde opdracht  **Een onderdeel van de geïntegreerde opdracht is een taalmodel afstemmen voor classificatie met data uit de beroepspraktijk.** |
| Week 8 | Bias detectie in grote taalmodellen – aanmaak synthetische/fake data  (Datapunten: Classification) | Voor bias detectie op sociale stigmatisering bij grote taalmodellen moeten data met prompts aangemaakt worden om via de techniek van masking prompts te genereren. De impact hiervan wordt geanalyseerd op emotion analysis op het IMDB film data set.  Theorie, oefeningen, oefening in functie van de geïntegreerde opdracht  **Voor de geïntegreerde opdracht wordt bias voor gender en age gedetecteerd in functie van automatische classificatie van CVs voor job screening. Hiervoor worden synthetische data voor gender, age, nationality aangemaakt.** |
| Week 9 | Bias detectie in grote taalmodellen – sentiment classificatie op synthetische data  (Datapunten: Classification) | Sentiment analyse toepassen op synthetische prompt data om negatieve en positieve stereotypering te kunnen analyseren.  Hiertoe wordt een mapping woorden lijst van positief en negatief geladen adjectieven gebruikt, en/of wordt een sentiment analyse tool die werd ontwikkeld tijdens de vorige colleges 'gerecycleerd'. Resulterende positieve en negatieve stereotypering wordt vergeleken voor verschillende taalmodellen.  **Voor de geïntegreerde opdracht** wordt bias voor gender en age gedetecteerd in functie van automatische classificatie van CVs voor job screening. Deze synthetische data met fake persoonlijke gegevens wordt automatisch toegevoegd aan open source CV data zonder persoonlijke gegevens. **De impact op classificate met en zonder deze fake persoonlijke gegevens wordt vergeleken voor verschillende taalmodellen.** |
| Week 10 | Bias detectie in grote taalmodellen – Bias verzachting – synthetische data  (Datapunten: Classification) | Op basis van sociale stigmatisering bias detectie resultaten, worden classificatie modellen opnieuw fine tuned d.m.v. nieuwe synthetische data met de bedoeling bias te verminderen. Dit wordt toegepast op emotie analyse met IMDB film data set.  Hiervoor worden tijdens dit college en deze fase synthetische data *gemaakt* via de techniek van masking.  Gebruik makend van TF-IDF wordt belangrijkste woord in input tekst gemaskeerd, toegepast op BBC text data set.  **Voor de geïntegreerde opdracht van CV classificatie, worden op basis van bias detectie resultaten classificatie modellen opnieuw fine tuned d.m.v. nieuwe synthetische data met de bedoeling bias te verminderen. Dit wordt toegepast op CV classificatie.** |
| Week 11 | Bias detectie in grote taalmodellen – Bias verzachting – synthetische data  (Datapunten: Classification) | Op basis van sociale stigmatisering bias detectie resultaten, worden classificatie modellen opnieuw fine tuned d.m.v. nieuwe synthetische data met de bedoeling bias te verminderen. Dit wordt toegepast op emotie analyse met IMDB film data set.  Hiervoor worden tijdens dit college en deze fase de synthetische data in vorig college gemaakt, *gebruikt* om ML model opnieuw te fine tunen en te evalueren  **Voor de geïntegreerde opdracht van CV classificatie, Op basis van bias detectie resultaten worden classificatie modellen opnieuw fine tuned d.m.v. nieuwe synthetische data met de bedoeling bias te verminderen. Dit wordt toegepast op CV classificatie.** |
| Week 12 | Text generation – text completion – Ngrams/conditional probability  (Datapunten: chatbot) | Textgeneration/completion met een traditionele, maar nog steeds actuele benadering, gebruik makend van ngrams, conditional probability  Theorie, oefeningen in functie van automatische sentence completion en spelling correctie, gebruik makend van Ngram dictionaries |
| Week 13 | Text generation – text completion – question-answering voor chatbot via fine tuning transformer model  (Datapunten: Chatbot) | Finetune een pretrained language model met data set (Stanford Question Answering Dataset - Squad) voor question answering.  Via overflowing tokens wordt omgegaan met data input die de maximum toegestane lengte van 512 (sub)wordtokens overschrijdt. |
| Week 14 | Text generation – text completion – question-answering voor chatbot via fine tuning transformer model  (Datapunten: Chatbot) | **Voor de geïntegreerde opdracht wordt op basis van de CV data automatisch een data set aangemaakt met context, question, en answer velden. Het CV data set bevat job profielen gecombineerd met fake data met naam, leeftijd en nationaliteit. Hiermee wordt een chatbot getraind die automatisch informatie geeft op een vraag over een bepaalde werknemer.** |
| Week 15 | Fine tuning NER using transformer model | NER fine tune transformer model for NER based on |

**Geïntegreerde opdracht beschrijving**

Gebaseerd op

[Ethical Programming Algorithms.pdf](https://hrnl.sharepoint.com/:b:/r/sites/CMI-DataScienceenAI/Shared%20Documents/Ontwikkelteam%20Algemeen/Jaar%203/semester%205/lesmateriaal/Ethical%20Programming%20Algorithms.pdf?csf=1&web=1&e=It4kSc)

* CV screening voor job assessment
* Profiel van een kandidaat : highlight automatisch belangrijkste eigenschappen van een kandidaat in zijn of haar CV
* Ranking van beste CVs voor een job profiel
* AI Blackbox challenge : welke specifieke data punten dragen bij tot selectie CV
* Bij selectieproces moet inclusiveness gerespecteerd worden: Bias detection voor leeftijd, gender, afkomst.

Gender, race, age, culturele bias wordt gedetecteerd voor sentiment classificatie van film reviews en **classificatie van CV's voor job screening**.

In eerste instantie wordt het IMDB dataset voor film review en CVs dataset geanalyseerd en gevisualisserd. Er wordt geverifieerd of er bias in de data steekt en of bijvoorbeeld bepaalde klasses over- of ondervertegenwoordigd zijn.

Bias kan ook voorkomen in open source taalmodellen zoals die onder de motorkap van een classifier zitten. Deze modellen worden hiervoor vergeleken om te beslissen welk model het best voor de classificatie of sentiment analyse opdracht wordt gebruikt. Hiervoor wordt de techniek van prompt engineering en masking toegepast. Daartoe wordt een synthetisch corpus of fake data gegenereerd om via prompt engineering en masking vooringenomenheid bloot te leggen. De prompts die via de synthetische data worden gegenereerd worden geëvalueerd op positieve of negatieve stereotypering via een sentiment classsifier en via een woordenlijst van positieve of negatieve adjectieven. Op die manier kan stereotypering voor gender, race, age, culturele bias voor meerdere taalmodellen worden vergeleken.

Voor het open source CV data set worden synthetische data met namen, nationaliteit en leeftijd geïnjecteerd in het open source CV data set dat geen persoonlijke gegevens bevat. **Gender, leeftijd, afkomst bias worden vergeleken voor verschillende taalmodellen**.

Om bias te verzachten kunnen samples aan een data set (film review, CVs) voor bepaalde klasses worden toegevoegd of weggenomen. Hiertoe moet opnieuw synthetische data worden aangemaakt. Dit kan bijvoorbeeld via een combinatie van masking, TF-IDF en cosine similarity om fake data aan te maken die zo dicht mogelijk bij echte data liggen.

Een model voor film review classificatie en CV classificatie wordt gefinetuned en vergeleken met en zonder gegenereerde fake data. Op die manier moet duidelijk worden of bias d.m.v. fake data wordt verzacht of niet.

Voor geselecteerde CVs wordt verder een tool ontwikkeld waar de belangrijkse eigenschappen worden highlighted, gebruik makend van SpaCy.

Eveneens wordt een *basic* **chatbot ontwikkeld waar via question answering** vragen kunnen gesteld worden over het profiel in het CV van een kandidaat.

Doorheen semester 5 worden de nodige vaardigheden voor de geïntegreerde opdracht aangebracht via kleinere opdrachten gelinkt aan o.a. het film review data set en andere open source data sets. De studenten moeten finaal de opdracht zelfstandig kunnen maken, gebruik makend van het data set met CVs voor job screening.

**Bronnen**

Jurafsky, D., & Martin, J. H. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.

<https://medium.com/>

<https://towardsdatascience.com/>

Rothman, D. (2022). *Transformers for Natural Language Processing: Build, train, and fine-tune deep neural network architectures for NLP with Python, Hugging Face, and OpenAI's GPT-3, ChatGPT, and GPT-4*. Packt Publishing Ltd.

Stevens, E., Antiga, L., & Viehmann, T. (2020). *Deep learning with PyTorch*. Manning Publications.

Ashish, V. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*, I.

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Vig, J. (2019). A multiscale visualization of attention in the transformer model. *arXiv preprint arXiv:1906.05714*.