SORBONNE UNIVERSITÉ

Early Detection of Psychotic Disorders: the Role of Emotions

Summer School on Digital Humanities
September 13TH TO September 14TH, UNAM, Mexico

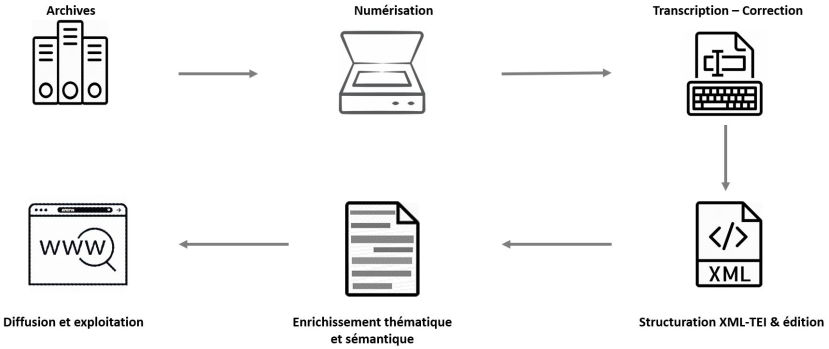
Motasem Alrahabi, ObTIC

Presentation

- 2010 : PhD in Linguistic Engineering from Sorbonne University, Paris.
- 2010 - 2018 : ICTE Lecturer - Sorbonne University UAE.
- 2018 - today: Research Engineer in Digital Humanities - ObTIC, Sorbonne Univ., Paris.
    - AI, NLP, Semantic and Discursive Analysis, Digital Publishing, ICT for Education.

L’Observatoire des textes, des idées et des corpus (ObTIC)

- ObTIC (former LabEx OBVIL) is a project team dedicated to Digital Humanities at Sorbonne University.
- Collaboration with the SCAI (Sorbonne Center for Artificial Intelligence) and the Datalab (National Library of France).
- ObTIC draws on the expertise acquired in:
  - Production and digital edition of data (see OBVIL Library).
  - Design and experimentation of text mining tools (TAL, AI, corpus analysis, textometry, etc.) for the human and social


Archives
Numérisation
Transcription – Correction
Structuration XML-TEI & édition
Enrichissement thématique et sémantique
Diffusion et exploitation
WWW
XML

Digital Publishing and Databases

- Automatic File Conversion Tool (Teinte)
- OBVIL Digital Library
- Valentin Haüy Digital Library (AVH)
- Sound heritage of poetry (ASP)
- Registers of the French Comedy (RCF-19)
- Revolutionary Opéra-comique Database (OCD)

UNIVERSITÉ D'ÉTÉ | BI-LICENCE "LETTRES – INFORMATIQUE" | ATELIERS
LETTRE D'INFORMATION | BLOG | FA
Observatoire de la vie littéraire
Rechercher...
OBVIL
ACTUALITÉS
PROJETS
BIBLIOTHÈQUE
CARNET DE RECHERCHE
DÉVELOPPEMENTS
PARTENAIRES INSTITUTIONNELS
RELATIONS INTERNATIONALES
Accueil Obvil > Bibliothèque
Critique
Molière
Mercure Galant
Apollinaire
Haine du th

Valentin HAÜY
Avec les aveugles et les malvoyants
Accueil
Archives
Auteurs
Recherche
Publications scientifiques
Contact

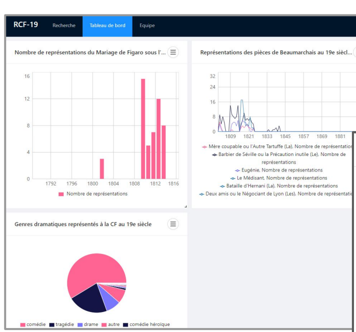
Démonstration de la future bibliothèque numérique Valentin Haüy

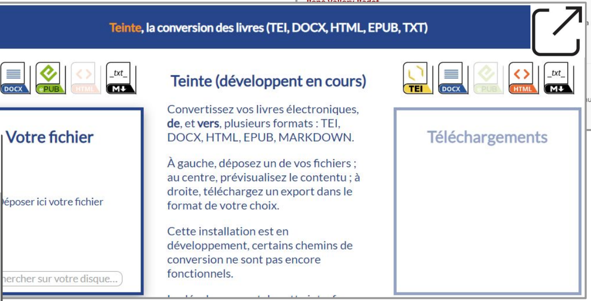
La bibliothèque patrimoniale Valentin Haüy est créée dans les années 1880 par Maurice de la Sizeranne. Ses fonds de 7000 ouvrages environ portent sur les aveugles et la cécité, et contiennent de nombreux textes uniques au monde.

L'association Valentin Haüy souhaite rappeler l'existence de cette histoire et préserver ces collections anciennes et fragiles. L'objectif du projet est donc de faire connaître l'histoire de la cécité et des aveugles aux premières personnes concernées et au grand public.

a même ville le 24 janvier 1933, gendre de Louis Pasteur dont

uteur dramatique, journaliste et voyageur français.


RCF-19
Recherche
Tableau de bord
Equipe
Nombre de représentations du Mariage de Figaro sous l'...
16
12
8
4
0
1792
1796
1800
1804
1808
1812
1816
Nombre de représentations
Représentations des pièces de Beaumarchais au 19e siècl...
32
24
16
8
0
1809
1821
1833
1845
1857
1869
1881
Mère coupable ou l'Autre Tartuffe (La), Nombre de représentation
Barbier de Séville ou la Précaution inutile (Le), Nombre de représentations
Eugénie, Nombre de représentations
Le Médisant, Nombre de représentations
Bataille d'Hernani (La), Nombre de représentations
Deux amis ou le Négociant de Lyon (Les), Nombre de représentati
Répartition des auteurs de la CF par genre
800
600
400
Genres dramatiques représentés à la CF au 19e siècle
comédie
tragédie
drame
autre
comédie héroïque



Teinte, la conversion des livres (TEI, DOCX, HTML, EPUB, TXT)

Teinte (développent en cours)

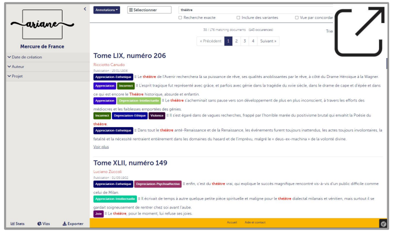
Convertissez vos livres électroniques, de, et vers, plusieurs formats : TEI, DOCX, HTML, EPUB, MARKDOWN.

À gauche, déposez un de vos fichiers ; au centre, prévisualisez le contenu ; à droite, téléchargez un export dans le format de votre choix.

Cette installation est en développement, certains chemins de conversion ne sont pas encore fonctionnels.

Text Exploration and Mining

- Obvie: Corpus Linguistic Analysis
- Elicom: Explore correspondence and letters
- Ariane: Semantic analysis of texts
- Tanagra: Mapping place names in texts
- Summarizer: Summarizing scientific articles
- Pandore: The toolbox for digital humanities



38 / 176 matching documents (143 occurences)

Tome LIX, numéro 206

Ricciotto Canudo

Publication : 15/01/1906

Appreciation-Esthetique ≡ Le théâtre de l'Avenir recherchera là sa puissance de rêve, ses qualités anoblissantes par le rêve, à côté du Drame Héroïque à la Wagner.

Appreciation Incorrect ≡ L'esprit tragique fut représenté avec grâce, et parfois avec génie dans la tragédie du xviie siècle, dans le drame de cape et d'épée et dans ce qui est encore le Théâtre historique, absurde et enfantin.

Appreciation Depreciation-Intellectuelle ≡ Le théâtre s'acheminait sans pause vers son développement de plus en plus inconscient, à travers les efforts des médiocres et les faiblesses emportées des génies.

Incorrect Depreciation-Ethique Violence ≡ Il s'est égaré dans de vagues recherches, frappé par l'horrible marée du positivisme brutal qui envahit la Poésie du théâtre.

Appreciation-Esthetique ≡ Dans tout le théâtre anté-Renaissance et de la Renaissance, les événements furent toujours inattendus, les actes toujours involontaires, la fatalité et la nécessité rentraient entièrement dans les domaines du hasard et de l'imprévu, malgré le « deus-ex-machina » de la volonté divine.

Voir plus

Tome XLII, numéro 149

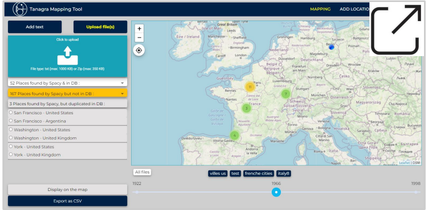
Luciano Zùccoli

Publication : 01/05/1902

Appreciation-Esthetique Depreciation-Psychoaffective ≡ enfin, c'est du théâtre vrai, qui explique le succès magnifique rencontré vis-à-vis d'un public difficile comme celui de Milan.

Appreciation-Intellectuelle ≡ Il écrivait de temps à autre quelque petite pièce spirituelle et maligne pour le théâtre dialectal milanais et vénitien, mais surtout il se gardait soigneusement de rentrer chez soi avant l'aube.

Joie ≡ Le théâtre, pour le moment, lui refuse ses joies.

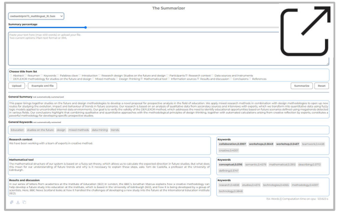


52 Places found by Spacy & in DB :

167 Places found by Spacy but not in DB :

3 Places found by Spacy, but duplicated in DB :

○ San Francisco - United States
○ San Francisco - Argentina
○ Washington - United States
○ Washington - United Kingdom
○ York - United States
○ York - United Kingdom



Paste your text here (max 400 words) or upload your file.
Two current options: Plain text format or XML

Choose title from list
□ Abstract □ Resumen □ Keywords □ Palabras clave □ Introduction □ Research design: Studies on the future and design □ Participants ☒ Research context □ Data sources and instruments □ DEFLEXOR methodology for studies on the future and design □ Mixed methods □ Design Thinking ☒ Mathematical tool □ Information sources ☒ Results and discussion □ Conclusions □ References

General Summary not automatically summarized

This paper brings together studies on the future and design methodologies to develop a novel proposal for prospective analysis in the field of education. We apply mixed research methods in combination with design methodologies to open up new routes for studying the evolution, impact and behaviour of trends in future scenarios. Our research is based on an analysis of qualitative data from secondary sources and interviews with experts, which we transform into quantitative data using fuzzy logic models applied to uncontrolled Internet data environments. Our goal is to verify the validity of the DEFLEXOR method, which addresses the need to identify educational opportunities based on future scenarios defined using megatrends detected in various fields. Our conclusions highlight that combining qualitative and quantitative approaches with the methodological principles of design thinking, together with automated calculations arising from creative reflection by experts, constitutes a powerful methodology for developing specific prospective studies.

General Keywords not automatically extracted

Education studies on the future design mixed methods data mining trends

Research context

We have been working with a team of experts in creative method.

Keywords: collaboration,0.5957 workshops,0.5649 workshop,0.5467 teamwork,0.4438 creative,0.4007

Mathematical tool

The mathematical structure of our system is based on a fuzzy set theory, which allows us to calculate the expected direction in future studies. But what does this mean for our understanding of future trends and why is it necessary to explain these steps, asks Tom de Castella, a professor at the University of Edinburgh.

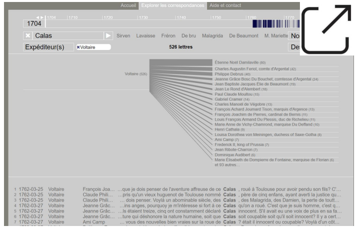
Keywords: conceptual,0.5116 semantic,0.4579 mathematical,0.3913 describing,0.3772 defining,0.3747

Results and discussion

In our series of letters from academics at the Institute of Education (IED) in London, the BBC's Jonathan Marcus explains how a creative methodology can help develop a future study into education at the institute, which is based in the University of Edinburgh (IED), and how it is being developed by a group of scientists. Here, BBC News Scotland looks at how it handled the challenges of developing a new study into the future at the International Education Institute (IED).

Keywords: research,0.4838 studies,0.4373 technologies,0.4065 methodology,0.4007 technology,0.3848

154 Words || Computation time on cpu: 123.623 s

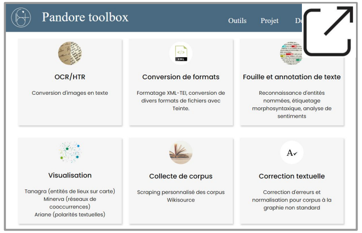


526 lettres

Voltaire (526)

Étienne Noël Damilaville (80)
Charles Augustin Feriol, comte d'Argental (42)
Philippe Debrus (40)
Jeanne Grâce Bosc Du Bouchet, comtesse d'Argental (24)
Jean Baptiste Jacques Élie de Beaumont (19)
Jean Le Rond d'Alembert (18)
Paul Claude Moultou (15)
Gabriel Cramer (14)
Charles Manoël de Végobre (13)
François Achard Joumard Tison, marquis d'Argence (13)
François Joachim de Pierres, cardinal de Bernis (11)
Louis François Armand Du Plessis, duc de Richelieu (11)
Marie Anne de Vichy-Chamrond, marquise Du Deffand (10)
Henri Cathala (9)
Louisa Dorothea von Meiningen, duchess of Saxe-Gotha (8)
Ami Camp (7)
Frederick II, king of Prussia (7)
Jean Ribote-Charron (7)
Dominique Audibert (6)
Marie Élisabeth de Dompierre de Fontaine, marquise de Florian (6)
et 93 autres…

```
1 1762-03-25 Voltaire  François Joa… …que je dois penser de l'aventure affreuse de ce Calas , roué à Toulouse pour avoir pendu son fils? C'…
2 1762-03-25 Voltaire  Claude Phili… …pris qu'un vieux huguenot de Toulouse nommé Calas , père de cinq enfans, ayant averti la justice qu…
3 1762-03-25 Voltaire  Claude Phili… … dois penser. Voylà un abominable siècle, des Calas , des Malagrida, des Damien, la perte de toutt…
4 1762-03-27 Voltaire  Jeanne Grâc… …ins anges, pourquoy je m'intéresse si fort à ce Calas qu'on a roué. C'est que je suis homme, c'est q…
5 1762-03-27 Voltaire  Jeanne Grâc… …ls étaient treize, cinq ont constamment déclaré Calas innocent. S'il avait eu une voix de plus en sa fa…
6 1762-03-27 Voltaire  Jeanne Grâc… …ture qui déshonore la nature humaine, soit que Calas soit coupable soit qu'il soit innocent? Il y a cert…
7 1762-03-27 Voltaire  Ami Camp     … vous des nouvelles bien vraies sur la roue de Calas ? était il innocent ou coupable? Voylà d'un côt…
```



OCR/HTR

Conversion d'images en texte

Conversion de formats

Formatage XML-TEI, conversion de divers formats de fichiers avec Teinte.

Fouille et annotation de texte

Reconnaissance d'entités nommées, étiquetage morphosyntaxique, analyse de sentiments

Visualisation

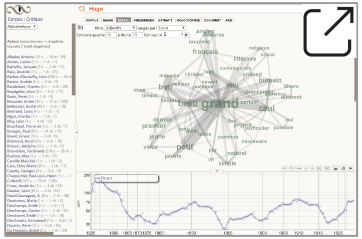
Tanagra (entités de lieux sur carte)
Minerva (réseaux de cooccurrences)
Ariane (polarités textuelles)

Collecte de corpus

Scraping personnalisé des corpus Wikisource

Correction textuelle

Correction d'erreurs et normalisation pour corpus à la graphie non standard



Auteur (occurrences — chapitres trouvés / total chapitres)

Albalat, Antoine (33 o. — 18 ch. / 69)
Arréat, Lucien (1 o. — 1 ch. / 1)
Bainville, Jacques (6 o. — 2 ch. / 10)
Baju, Anatole (1 o. — 1 ch. / 21)
Barbey d'Aurevilly, Jules (336 o. — 46 ch
Barine, Arvède (2 o. — 2 ch. / 8)
Baudelaire, Charles (3 o. — 3 ch. / 29)
Bazalgette, Léon (6 o. — 5 ch. / 17)
Bazin, René (3 o. — 1 ch. / 5)
Beaunier, André (45 o. — 17 ch. / 126)
Bellessort, André (44 o. — 4 ch. / 10)
Bertrand, Louis (4 o. — 1 ch. / 1)
Bigot, Charles (1 o. — 1 ch. / 1)
Bloy, Léon (4 o. — 4 ch. / 32)
Bouchaud, Pierre de (3 o. — 2 ch. / 5)
Bourget, Paul (82 o. — 24 ch. / 70)
Bovet, Ernest (10 o. — 3 ch. / 6)
Bremond, Henri (3 o. — 2 ch. / 16)
Brisson, Adolphe (10 o. — 1 ch. / 2)
Brunetière, Ferdinand (218 o. — 35 ch. /
Buchon, Max (3 o. — 3 ch. / 10)
Camille Mauclair (1 o. — 1 ch. / 2)
Caro, Elme-Marie (28 o. — 2 ch. / 17)
Casella, Georges (3 o. — 3 ch. / 9)
Charpentier, Paul-Louis-Henri (1 o. — 1
Collectif (127 o. — 75 ch. / 166)
Croze, Austin de (7 o. — 4 ch. / 34)
Daudet, Léon (52 o. — 8 ch. / 57)
David-Sauvageot, A. (9 o. — 7 ch. / 48)
Deraismes, Maria (1 o. — 1 ch. / 2)
Deschamps, Émile (1 o. — 1 ch. / 1)
Deschamps, Gaston (2 o. — 2 ch. / 55)
Deschanel, Émile (1 o. — 1 ch. / 19)
Des Essarts, Emmanuel (7 o. — 4 ch. / 1
Doumic, René (37 o. — 8 ch. / 48)
Du Fresnois, André (2 o. — 1 ch. / 24)

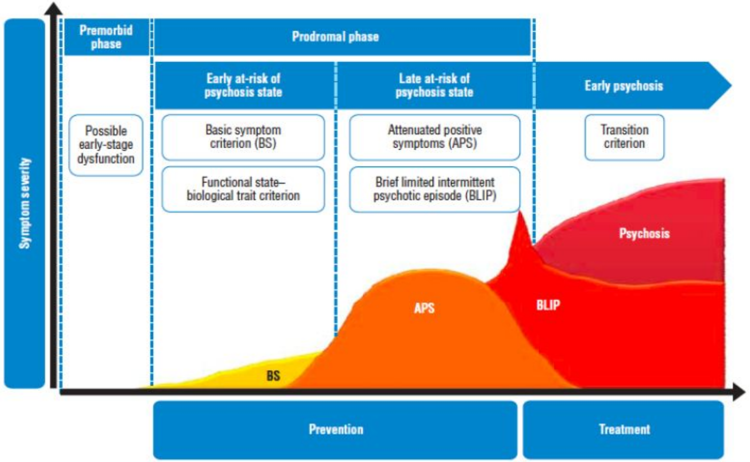
APAISE Project

Context

- Project APAISE: *Apprentissage Profond pour l'Analyse Informatisée de la Subjectivité et des Émotions dans les troubles psychotiques émergents*.
  - Deep Learning for Computational Analysis of Subjectivity and Emotions in Emerging Psychotic Disorders
- Funding: Fondation de France (Psychiatric Disease Research program)
  - Start: September 2023
- Involved teams:
  - ObTIC - Sorbonne University: Motasem Alrahabi, Jean Marie Tshimula (soon)
  - INSERM: Marie-Odile Krebs, Julien Descles, Valeria Lucarini
  - Brest University Hospital Centre: Michel Walter, Christophe Lemey, Deok-Hee Kim-Dufor
- Goals:
  - Early detection of psychotic disorders: the role of emotions (and the retelling of memories).
  - Better management of patients at risk and slowing down their evolution towards chronicity.

State of the Art

- There are numerous NLP studies that analyse *language* in the emerging psychotics disorders.
- Identify discriminating anomalies to predict the evolution of patients with differents methods, on different levels: prosody, syntax, semantics, vocabulary, discourse, dialogue…


Symptom severity
Premorbid phase
Prodromal phase
Early at-risk of psychosis state
Late at-risk of psychosis state
Early psychosis
Possible early-stage dysfunction
Basic symptom criterion (BS)
Functional state–biological trait criterion
Attenuated positive symptoms (APS)
Brief limited intermittent psychotic episode (BLIP)
Transition criterion
Psychosis
APS
BLIP
BS
Prevention
Treatment

- Clinical evolution of psychosis: 3 phases
    - At-risk, Not-at-risk, Psychotic "CAARMS" score [Yung et al., 2005]

- Challenge: reduce the prodromal phase (at-risk) before the chronic phase.

→ [Magaud et al., 2010], [Tanguy et al., 2011], [Register-Brown, Hong LE 2014], [Bedi et al. 2015], [Bazziconi, 2018], [Ratana et al., 2019], [Lucarini et al., 2023]...

Working Hypothesis

- Clinical studies: the subjective dimension of language expression (emotions, sentiments, perceptions…) could reflect in the patient's speech a disturbed relationship to the world and to themself.
- Few academic works in this field [Tshimula et al., 2022], [Saffar, 2023].
- Hypothesis: the analysis of subjective modalities could play an important role in the early detection of psychosis (during the prodromal phase).

→ Classification problem: given a labeled text as input, what would its class belong to?

1	**At-risk (A)**
2	**Not-at-risk (N)**
3	**Psychotic (P)**
4	**Control (C)**

Data Preprocessing

Data

- Our corpus consists of rare psychiatric interviews:
  - Open dialogues between psychiatrists and patients (15 to 30 years old).
- About 250 audio interviews (differents patients).
  - Currently: 134 interviews (≈ 1 million tokens).
- Patient speech is characterised by:
  - Verbal pauses and disfluencies, hesitations, silences...
  - Disorganised speech (tangential and incoherent), broken syntax…
  - Low lexical density, short sentences...
  - Particular use of personal pronouns (moi, je, me, mon, ma, mes)
  - Emotions: stress, anger, violence, euphoria, joy, suffering…

1	**At-risk (A)**	**65 texts**
2	**Not-at-risk (N)**	**17 texts**
3	**Psychotic (P)**	**19 texts**
4	**Control (C)**	**33 texts**

Data Preprocessing

- We conducted a series of preprocessing on the data:
  - Manual transcription of audio files into text format.
  - Anonymisation: identification of named entities with SpaCy, then manual correction.
  - Oral errors are not corrected: unfinished words, morpho-syntax errors, conjugations...
  - Keep verbal disfluencies and hesitations: *ah, euh, hum, hmm, hein, ben, bah, pfff*…

Data labeling

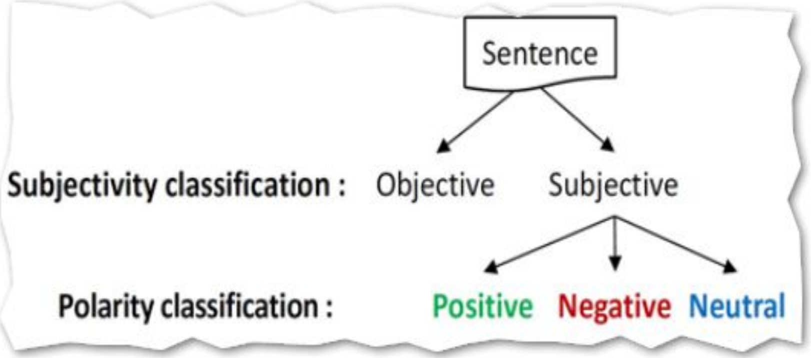
- Only for patients' speech
- Applied features for each patient (text level analysis):
  - Label #1 → Average of sentence length
  - Label #2 → Average of the personal pronouns
  - Label #3 → Average of the verbal disfluencies
  - Label #4 → Lexical density of vocabulary
  - Label #5 → Subjective modalities: emotions, sentiments, opinions…

Subjective Modalities

Data labeling: Subjective Modalities

- Allows us to capture subjective content:
  - Polarity: positive, negative, neutral or mixed.
  - Source and target.
  - Intensity (normal, strong, etc.).
  - Aspects of the analyzed object.

→ [Turney, 2002]; [Wiebe et al., 2005]; [Pang and Lee, 2008] [Balahur et al., 2011]; [Zhang and Bing 2017]…



- Need for more fine-grained classification:
  - GoEmotions: 27 labels for emotions in English [Demszky et al., 2020].

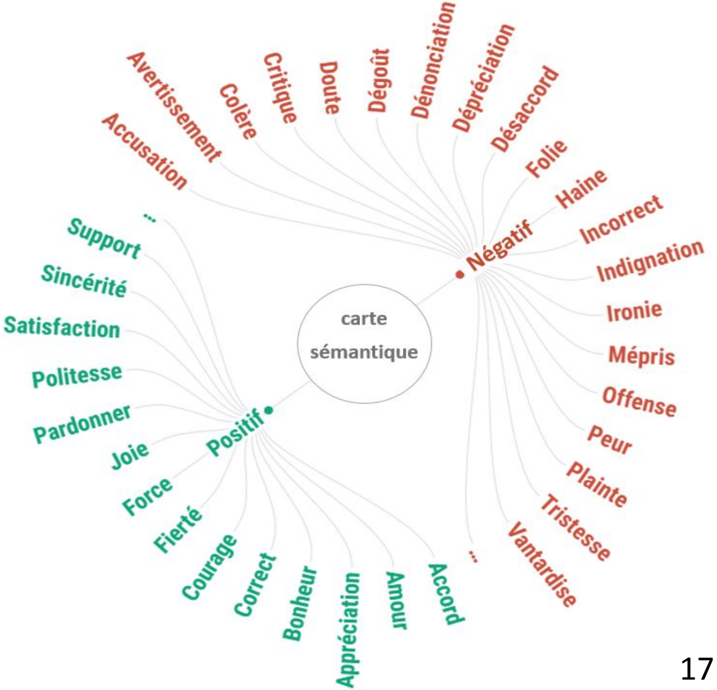
Data labeling: Subjective Modalities

- Examples from the dataset (approximate translation):
  - **One time I took my Swiss army knife and went for a walk and I, I just wanted to shove it down my throat.**
    - *Une fois j'avais pris mon couteau suisse et j'étais parti me promener et je, je voulais juste me le planter dans la gorge.* [Violence, Patient 2]
  - **I couldn't tell the difference between, between if I was in a dream or if I was in reality.**
    - *J’arrivais plus à faire la différence entre, entre si j’étais dans un rêve ou si j’étais dans la réalité.* [Hallucination, Patient 15]
  - **I just want to drink, until, finally, [I lose reason], because I have, I am in control of myself all the time.**
    - *J'ai juste envie de boire, jusqu'à, enfin, la déraison, enfin parce que j'ai, je suis tout le temps dans le contrôle de moi-même.* [Addiction, Patient 401]
  - **Each time I thought I had zeros [in exams], I realized that in fact uh they were just kidding me.**
    - *A chaque fois je pensais avoir des zéros, je me rendais compte qu'en fait euh on se foutait juste de ma gueule* [Mockery, Patient 1101]

Data labeling: Subjective Modalities

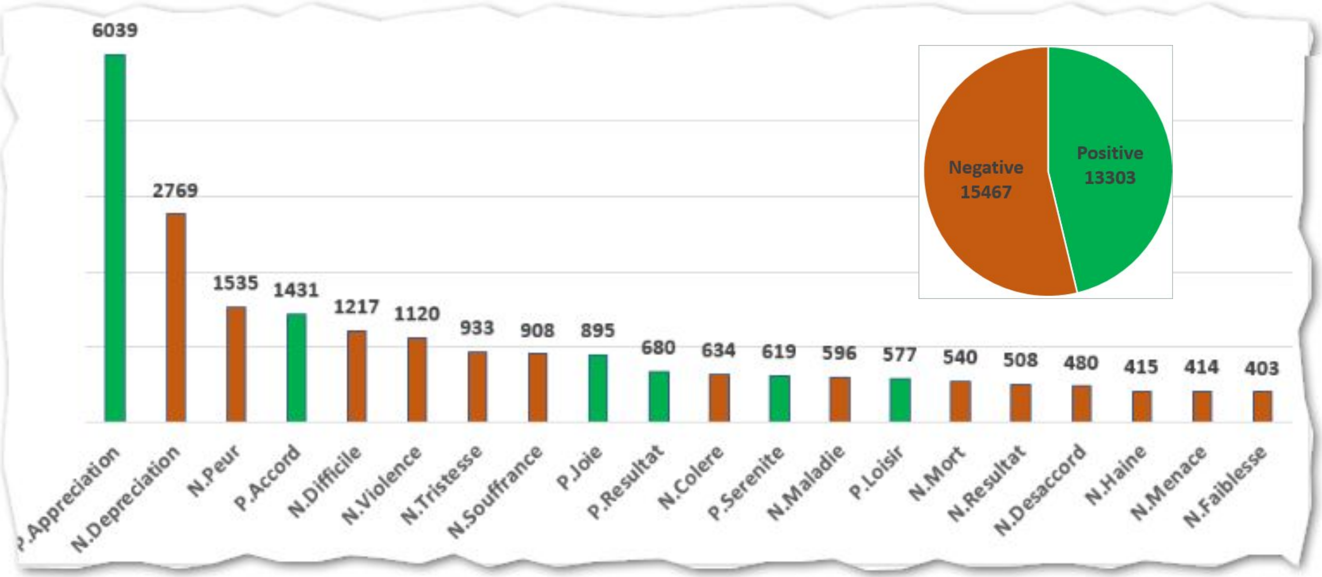
- Linguistic Ontology [Alrahabi, 2016, 2021]
  - ≈ 3500 observable markers (patterns)
  - Lexical categories: verbs, adjectives, adverbs, phrases…

- Fine-grained annotations:
  - Classified first as positive, negative or neutral.
  - Grouped into 82 sub-categories: anxiety, stress, anger, violence, joy, suffering…
  - Adaptation to the current project: set up new categories, consider oral speech, existing oral pronunciation errors…



Data labeling: Subjective Modalities

- Lexicon-based annotation tool (Textolab):
  - 28770 annotations / 134 texts / 1,050,144 words.



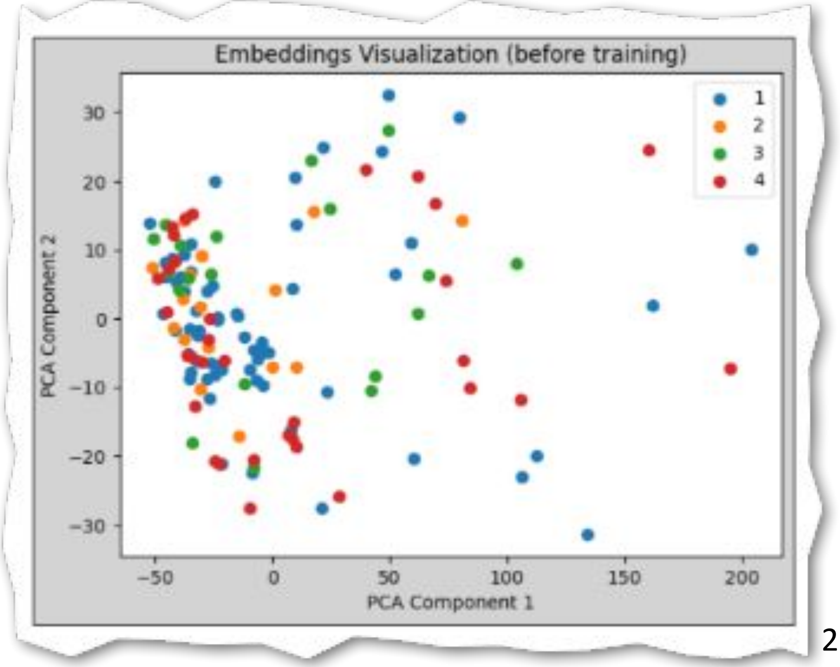
Annotation results can be consulted via a web interface (Ariane)

Supervised Classification

Data Representation (Embeddings)

- Vector representation: Camembert LM (https://huggingface.co/camembert-base)
- Embeddings are created using the "CamembertTokenizer" (based on "WordPiece").
- Embeddings are associated with features:
  - 4 linguistic labels
  - 82 emotion labels

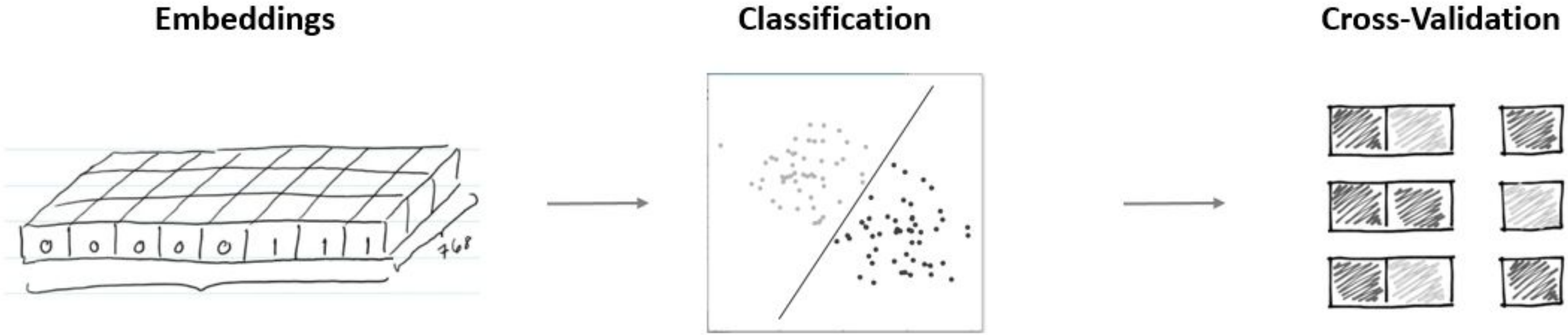
    → Dimensions: 768 + 86 = 854



- Embeddings visualized with PCA:
  - No underlying clusters.

Supervised classification

- Use of traditional machine learning algorithms (no enough data for Deep Learning)



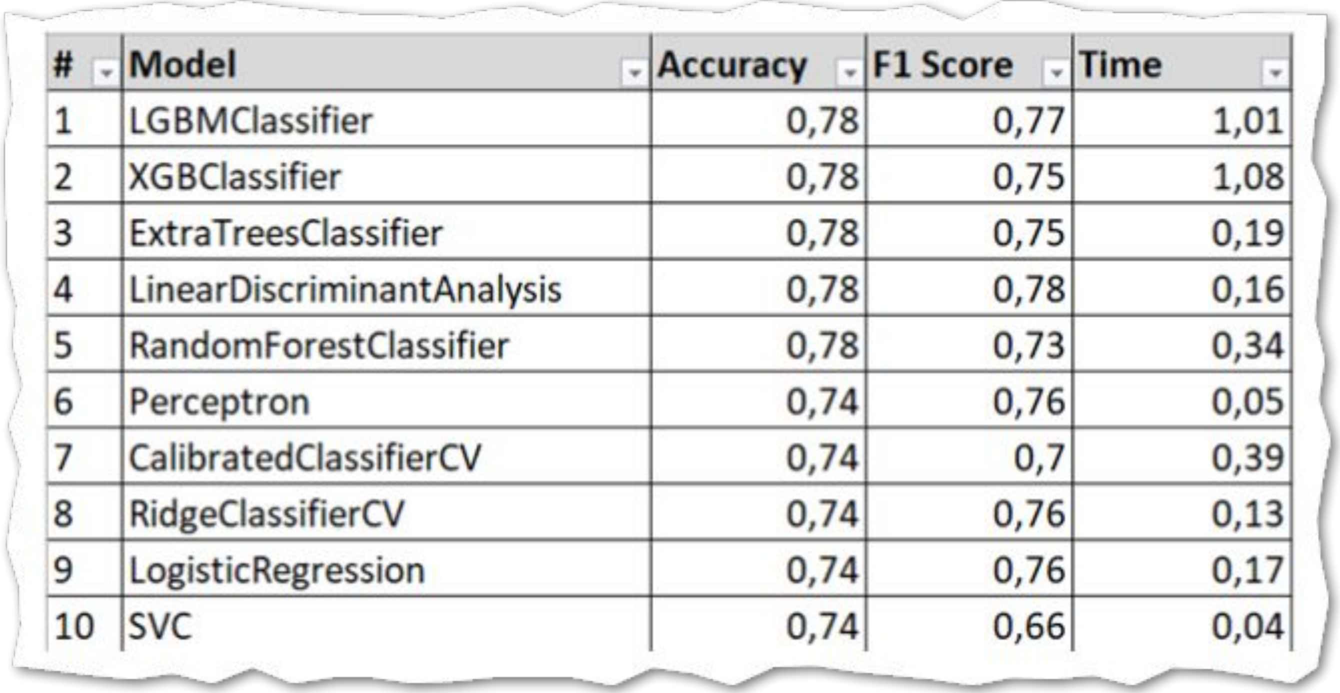
Embeddings

Classification

Cross-Validation

Evaluation of classification models

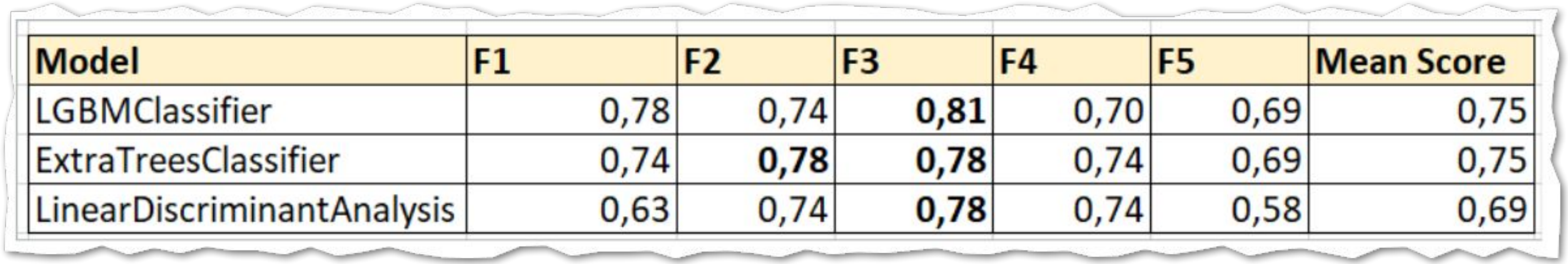
- LazyPredict library (https://pypi.org/project/lazypredict/)



#	Model	Accuracy	F1 Score	Time
1	LGBMClassifier	0,78	0,77	1,01
2	XGBClassifier	0,78	0,75	1,08
3	ExtraTreesClassifier	0,78	0,75	0,19
4	LinearDiscriminantAnalysis	0,78	0,78	0,16
5	RandomForestClassifier	0,78	0,73	0,34
6	Perceptron	0,74	0,76	0,05
7	CalibratedClassifierCV	0,74	0,7	0,39
8	RidgeClassifierCV	0,74	0,76	0,13
9	LogisticRegression	0,74	0,76	0,17
10	SVC	0,74	0,66	0,04

Evaluation of the best classification models

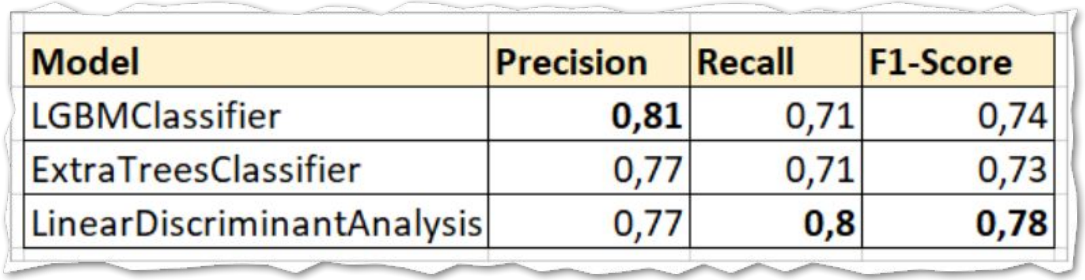
- Perform a cross-validation in terms of accuracy:
  - assessing model performance
  - tuning hyperparameters
  - ensuring generalization to new data
  - etc.



Model	F1	F2	F3	F4	F5	Mean Score
LGBMClassifier	0,78	0,74	**0,81**	0,70	0,69	0,75
ExtraTreesClassifier	0,74	**0,78**	**0,78**	0,74	0,69	0,75
LinearDiscriminantAnalysis	0,63	0,74	**0,78**	0,74	0,58	0,69

Evaluation of the best classification models

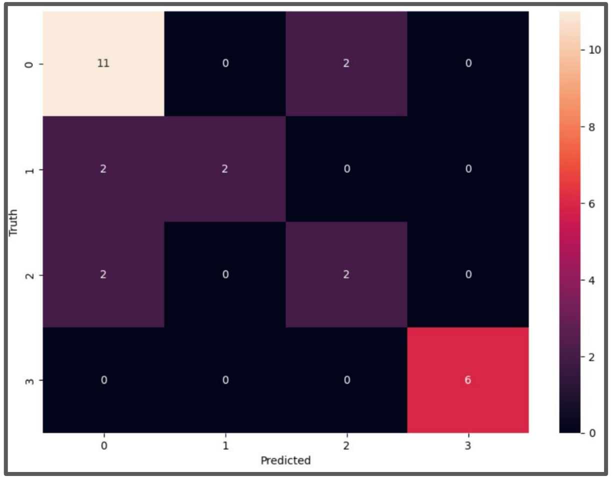
- Precision, recall and f-score:
  - evaluate the performance of classification models (quality, completeness and overall performance)
  - offer interpretable measures of a model's performance
  - aid in choosing the most appropriate machine learning algorithm for a given problem
  - etc.



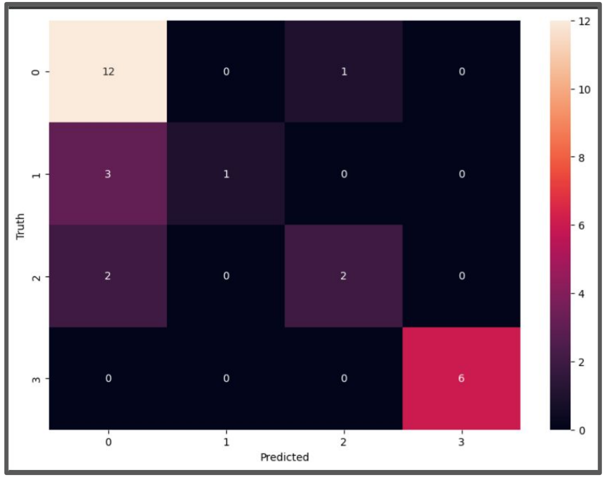
Model	Precision	Recall	F1-Score
LGBMClassifier	**0,81**	0,71	0,74
ExtraTreesClassifier	0,77	0,71	0,73
LinearDiscriminantAnalysis	0,77	**0,8**	**0,78**

Evaluation of the best classification models

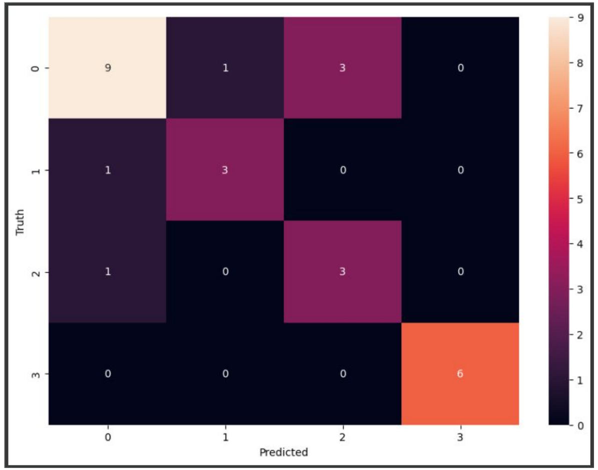
- Confusion Matrix:
  - offers a general overview of the performance of a classification model by summarizing the counts of true positive, true negative, false positive, and false negative predictions



LGBMClassifier

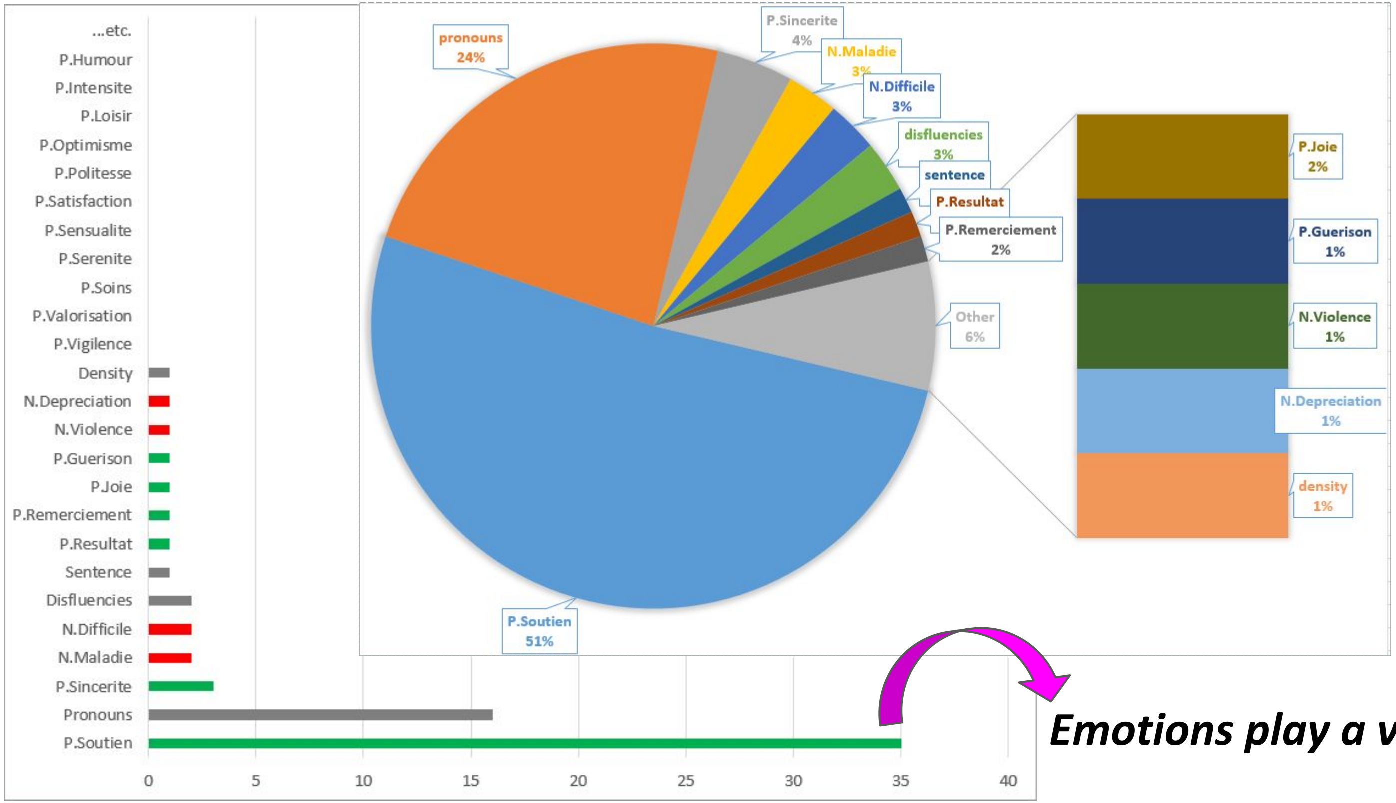


ExtraTreesClassifier



LDA

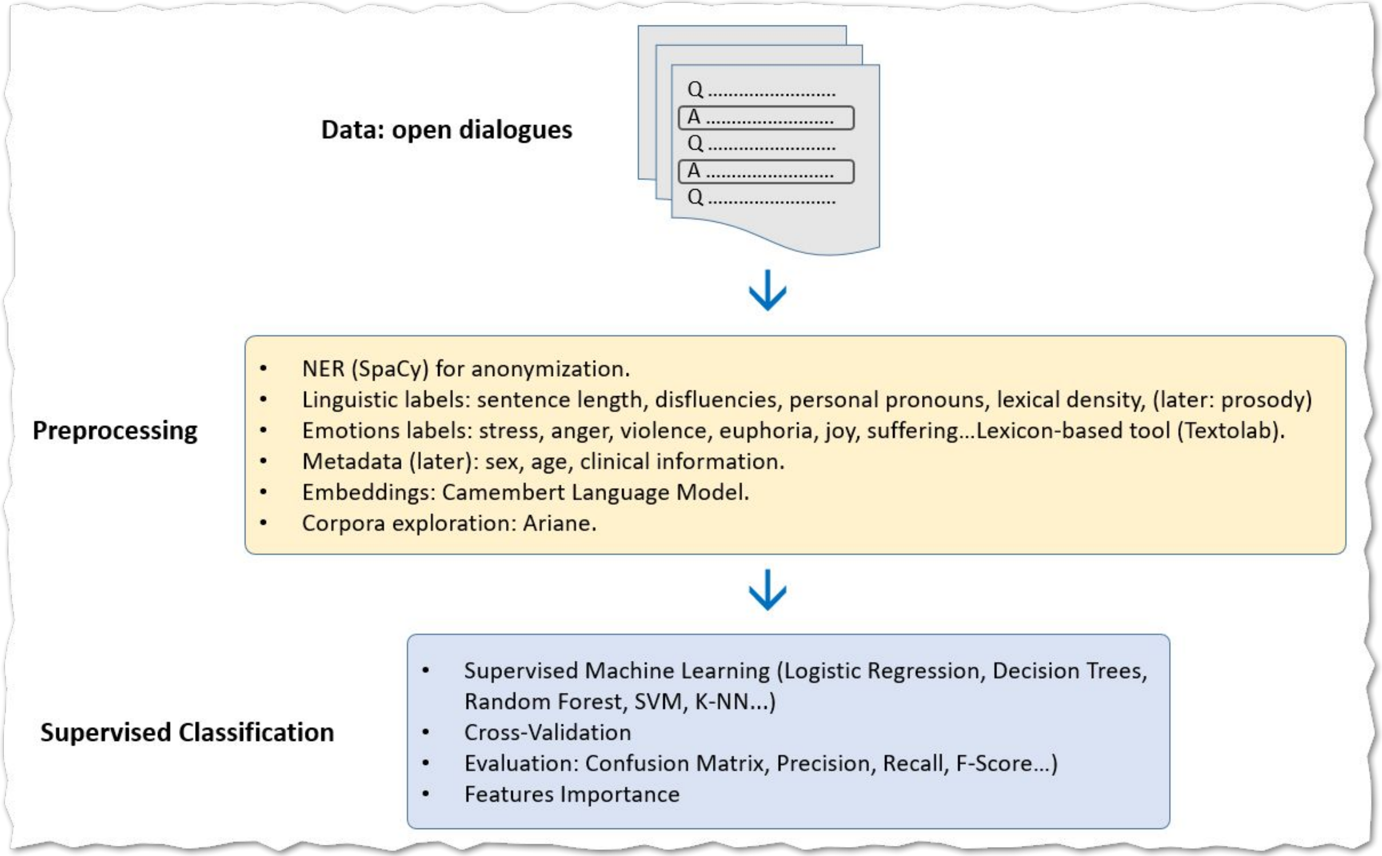
Analysing the importance of labels (features)



Emotions play a very important role

Conclusion

Synthesis: Hybrid Approach



Data: open dialogues

Preprocessing

- NER (SpaCy) for anonymization.
- Linguistic labels: sentence length, disfluencies, personal pronouns, lexical density, (later: prosody)
- Emotions labels: stress, anger, violence, euphoria, joy, suffering…Lexicon-based tool (Textolab).
- Metadata (later): sex, age, clinical information.
- Embeddings: Camembert Language Model.
- Corpora exploration: Ariane.

Supervised Classification

- Supervised Machine Learning (Logistic Regression, Decision Trees, Random Forest, SVM, K-NN…)
- Cross-Validation
- Evaluation: Confusion Matrix, Precision, Recall, F-Score…)
- Features Importance

Preliminary Results and Perspectives

- Recruitment of a postdoctoral fellow (oct. 2023).
- Use data sampling or sliding window techniques.
- Create a multi-label emotion model [Tao et al. 2020], [Demszky et al., 2020].
- Cross with other information:
  - Prosodic analysis: measurement of silence and intonation, in progress with INSERM [Lucarini et al., 2023].
  - Metadata: gender, age, clinical observations (risk of psychosis, consumption of products, etc.).

Thank you for your attention !

Bibliographic references

- Alexandre D, Alrahabi M, Gay F., Riguet M. “Le médical et le social: analyse sémantique des rapports de l’immersion d’étudiants de médecine dans le Samu social”, in Humanités Numériques Littéraires, sous la dir. de Didier Alexandre, Paris, Classiques Garnier, 2021
- Alrahabi M, Ariane: dispositif de fouille et de lecture synthétique de textes, Actes de DigitAl Humanities and cuLtural herItAge: data and knowledge management and analysis, Jan 2021, Montpellier, France.
- Bedi G, Carrillo F, Cecchi GA, Slezak DF, Sigman M, Mota NB, Ribeiro S, Javitt DC, Copelli M, Corcoran CM. Automated analysis of free speech predicts psychosis onset in high-risk youths. NPJ Schizophr. 2015.
- Bazziconi P-F, 2018, thèse de doctorat de médecine DOI: 10.13140/RG.2.2.16573.82400, Advisor: Christophe Lemey; Michel Walter
- Bazziconi P-F, Bleton L, et al.. L’Information Psychiatrique, 2019/2 (95), 89-94.
- Demszky, Dorottya & Movshovitz-Attias, Dana & Ko, Jeongwoo & Cowen, Alan & Nemade, Gaurav & Ravi, Sujith. (2020). GoEmotions: A Dataset of Fine-Grained Emotions. 4040-4054. 10.18653/v1/2020.acl-main.372.
- Lucarini et al. 2021. “Conversational metrics, psychopathological dimensions and self-disturbances in patients with schizophrenia” European Archives of Psychiatry and Clinical Neuroscience.
- Register-Brown K, Hong LE. Schizophrenia research. 2014 Dec; 160(1-3):20-6.
- Tao, J., Fang, X. Toward multi-label sentiment analysis: a transfer learning based approach. J Big Data 7, 1 (2020).
- Yung AR, Yuen HP, McGorry PD, Phillips LJ, Kelly D, Dell'Olio M, Francey SM, Cosgrave EM, Killackey E, Stanford C, Godfrey K, Buckby J. Mapping the onset of psychosis: the Comprehensive Assessment of At-Risk Mental States. Aust N Z J Psychiatry. 2005