# Article information:

From Decision Trees to Transformers: Comparing Sentiment Analysis Models for Macedonian Restaurant Reviews | by Danilo Najkov | Mar, 2023 | Towards Data Science
<https://towardsdatascience.com/from-decision-trees-to-transformers-comparing-sentiment-analysis-models-for-macedonian-restaurant-4c2d931ec021>

# Article summary:

1. Les modèles de traitement du langage naturel ont traditionnellement été développés pour les langues populaires telles que l'anglais et l'espagnol, mais avec la montée en puissance du commerce électronique, même les langues moins courantes comme la macédonienne génèrent des quantités importantes de données grâce aux avis en ligne.

2. Dans cet article, l'auteur explore et compare différents modèles d'analyse de sentiment pour les avis de restaurants macédoniens, allant des arbres de décision classiques aux techniques d'apprentissage profond modernes et aux transformateurs.

3. Les modèles utilisés comprennent LASER embeddings, Multilingual Universal Sentence Encoder et OpenAI Ada v2, ainsi que des algorithmes tels que Random Forests, XGBoost, Support Vector Machines et Deep Learning.

# Article rating:

Appears moderately imbalanced: The article provides some useful information, but is missing several important points or pieces of evidence that would be required to present the discussed topics in a balanced and reliable way. You are encouraged to seek a more balanced perspective on the presented issues by exploring the provided research topics and looking at different information sources.

# Article analysis:

L'article "From Decision Trees to Transformers: Comparing Sentiment Analysis Models for Macedonian Restaurant Reviews" publié sur Towards Data Science en mars 2023, présente une étude comparative de différents modèles d'analyse de sentiment pour les avis de restaurants en macédonien. L'auteur souligne que les langues moins courantes comme le macédonien ont vu peu de développement dans le domaine du traitement automatique du langage naturel (TALN), mais que la pandémie COVID-19 a généré une quantité importante de données en ligne pour ces langues. L'étude vise à aider les entreprises à mieux comprendre le sentiment des clients et à améliorer leurs services.

Le contenu de l'article est bien structuré et commence par expliquer la préparation des données, y compris la conversion des textes cyrilliques et latins, ainsi que la suppression des émojis. Ensuite, l'auteur décrit trois modèles d'embedding : LASER, Multilingual Universal Sentence Encoder (MUSE) et OpenAI Ada v2. Il explique comment créer les embeddings avec chaque modèle et fournit des exemples de code Python pour chacun.

Ensuite, l'auteur compare plusieurs modèles d'apprentissage automatique pour prédire le sentiment dans les avis de restaurants en macédonien. Les modèles incluent Random Forests, XGBoost, Support Vector Machines (SVM) et Deep Learning. Pour chaque modèle, l'auteur fournit une brève explication théorique ainsi qu'un exemple de code Python pour entraîner et tester le modèle.

Cependant, il y a quelques biais potentiels dans cet article. Tout d'abord, l'auteur ne mentionne pas si les résultats obtenus sont significatifs ou non. De plus, il n'y a pas d'évaluation quantitative ou qualitative approfondie des performances des différents modèles présentés. En outre, l'article ne discute pas suffisamment les limites potentielles liées aux données utilisées ou aux méthodes employées.

Enfin, il convient également de noter que l'article peut être considéré comme promotionnel car il utilise plusieurs bibliothèques open-source telles que LASER et MUSE qui sont développées par Facebook AI Research et Google AI respectivement.

Dans l'ensemble, cet article fournit un bon aperçu des différentes techniques utilisées pour analyser le sentiment dans les avis de restaurants en macédonien. Cependant, il manque une analyse critique plus approfondie des résultats obtenus ainsi qu'une discussion sur les limites potentielles liées aux données utilisées ou aux méthodes employées.

# Topics for further research:

* Limitations of sentiment analysis models for low-resource languages
* Evaluation metrics for sentiment analysis models
* Challenges in preprocessing text data for sentiment analysis
* Comparison of traditional machine learning models and deep learning models for sentiment analysis
* Importance of domain-specific training data for sentiment analysis
* Ethical considerations in using sentiment analysis for customer feedback analysis.

# Report location:

<https://www.fullpicture.app/item/536ffc3379278b89043e9310b97a339b>