



MELUSINE: TRAITEMENT AUTOMATIQUE DES EMAILS

Constantin Van de Velde | Benoît Lebreton | Florian Arthur + tous les contributeurs
Data Scientist Data Scientist Data Scientist

Quantmetry
Part of Capgemini Invent



initié par :  MAIF
assureur militant

1 MOTIVATIONS

- Le pré-traitement d'emails et leurs routages aux bonnes boîtes aux lettres peut-être **long et fastidieux**. **Particuliers** et **professionnels** aimeraient avoir une solution à la fois « **clé en main** » et **modulable** pour pouvoir traiter leurs emails plus rapidement.
- Plus particulièrement, dans le **secteur de l'assurance**, les entreprises reçoivent un nombre considérable d'e-mails avec différents sujets et degrés d'importance. **Mieux comprendre** le contenu d'un e-mail permettra d'améliorer la **qualité de la réponse**, d'optimiser le temps de réponse et d'offrir une meilleure expérience client.

2 OBJECTIFS

La communauté Open Source comme levier d'innovation

La MAIF a mis à contribution la **communauté Open-Source** pour proposer une **bibliothèque Python** permettant la classification des e-mails et l'extraction d'informations clés.

Melusine: une bibliothèque Python de haut niveau

Melusine est une bibliothèque Python de haut niveau pour la **classification des e-mails** et l'extraction de mots clés, écrite en Python et capable de s'exécuter sur Scikit-Learn, Tensorflow 2 et Keras. Elle met l'accent sur les e-mails rédigés en français.

3 MÉTHODOLOGIE

Le **traitement** des emails nécessite plusieurs **phases de prétraitement** comme le nettoyage des mots non importants pour les algorithmes. Le package embarque un subpackage ``prepare_email`` qui permet de faire le nettoyage textuel nécessaire.

Dans un contexte d'efficacité pour une entreprise, on pourrait imaginer un **affichage des mots-clés** d'un email qui soit globalement long à lire, cela est possible grâce au subpackage ``summarizer`` qui permet d'extraire les mots clés d'un texte.

Le subpackage ``models`` est utilisé pour **classer les e-mails** selon des catégories prédéfinies par l'utilisateur ou calculer le score de sentiment basé sur le sentiment décrit par l'utilisateur avec des mots clés.

Le subpackage ``ethics_guidelines`` fournit un guide d'éthique pour évaluer le projet d'IA, avec des lignes directrices et un questionnaire.

Le questionnaire est basé sur des critères issus notamment des **travaux de la Commission européenne** et regroupés par catégories.

Enfin, Melusine intègre différents **connecteurs** pour se brancher à des systèmes existants, et se basent sur des outils permettant son **passage en production** et sa mise à l'échelle sur des millions de mails: Melusine est **'Scalable'**.

4 MELUSINE OPEN SOURCE

CHIFFRES CLÉ (GITHUB MELUSINE)

60 releases dont une majeure avec BERT

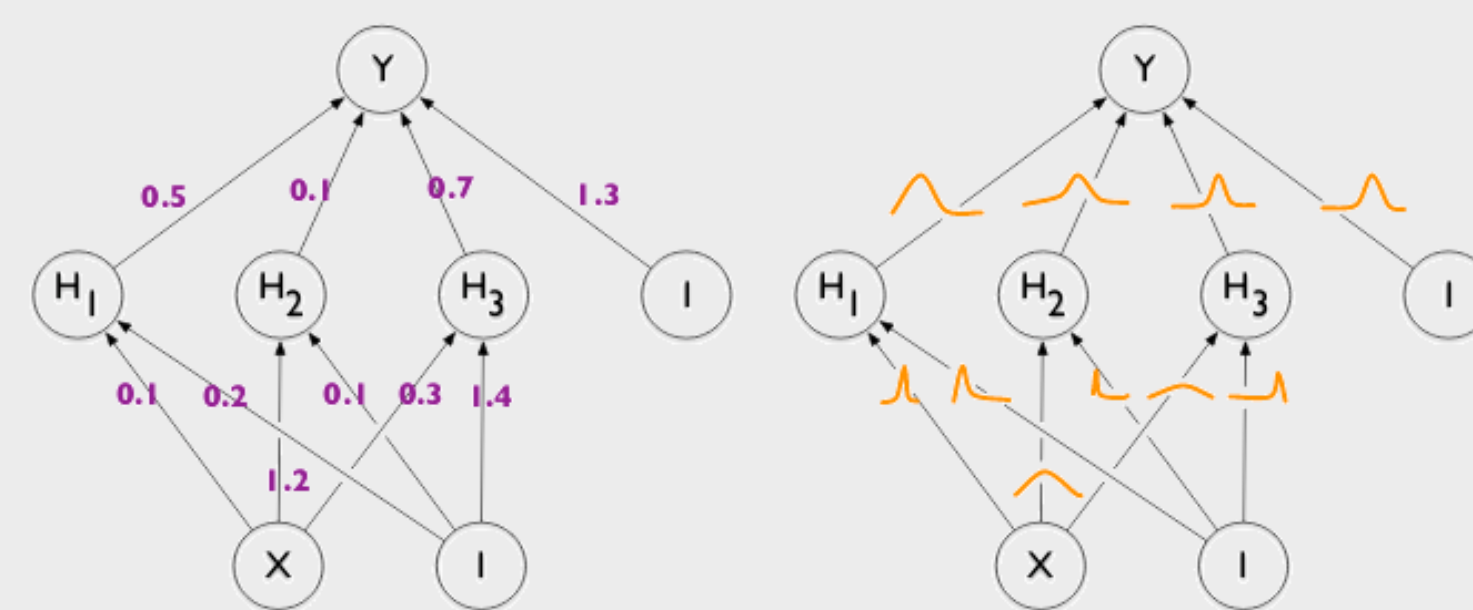
+ de **300 stars**  Starred 306


35 issues résolues

+ de **20 contributeurs** au projet dont plusieurs entreprises:
Quantmetry part of Capgemini Invent
Verlingue
Société Générale ...
... et bien entendu MAIF

DERNIÈRES ACTUALITÉS – MELUSINE AVEC ESTIMATION D'INCERTITUDES

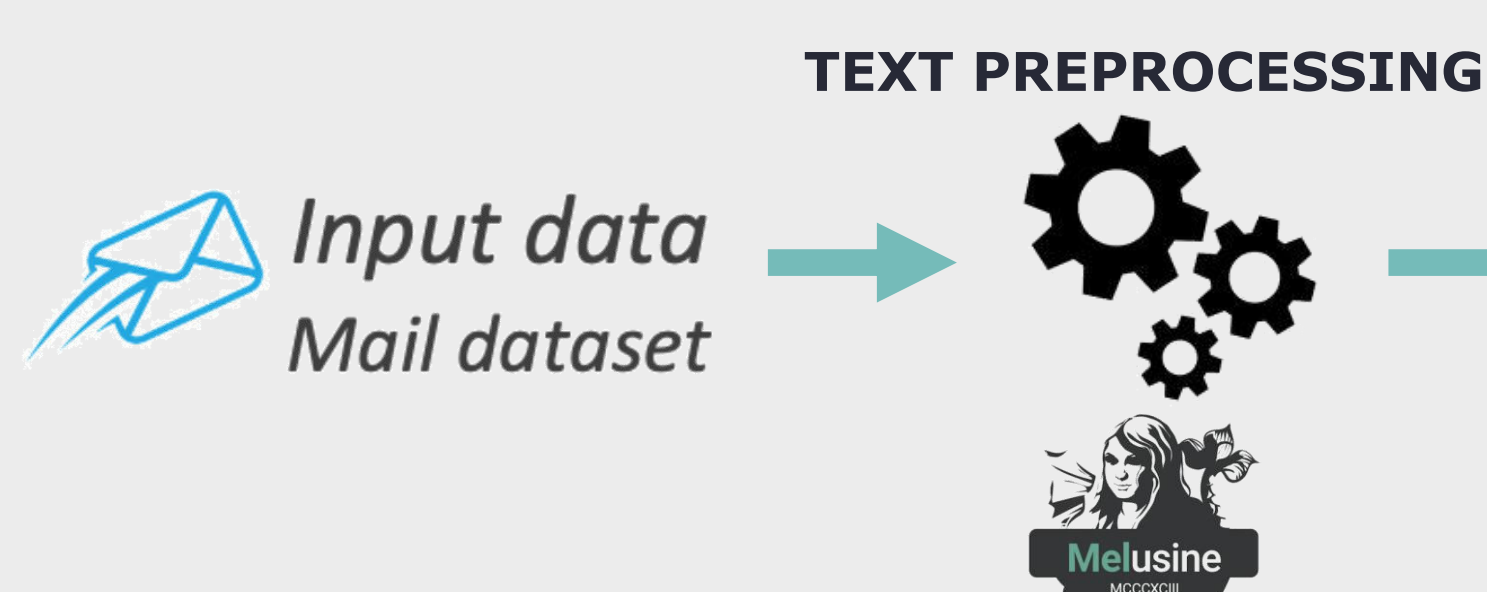
Fig. 1.: Exemple d'un réseau de neurone avec une distribution de probabilité au lieu d'un scalaire pour les étapes de propagation, permettant d'avoir en sortie une distribution et ainsi une estimation d'incertitude.



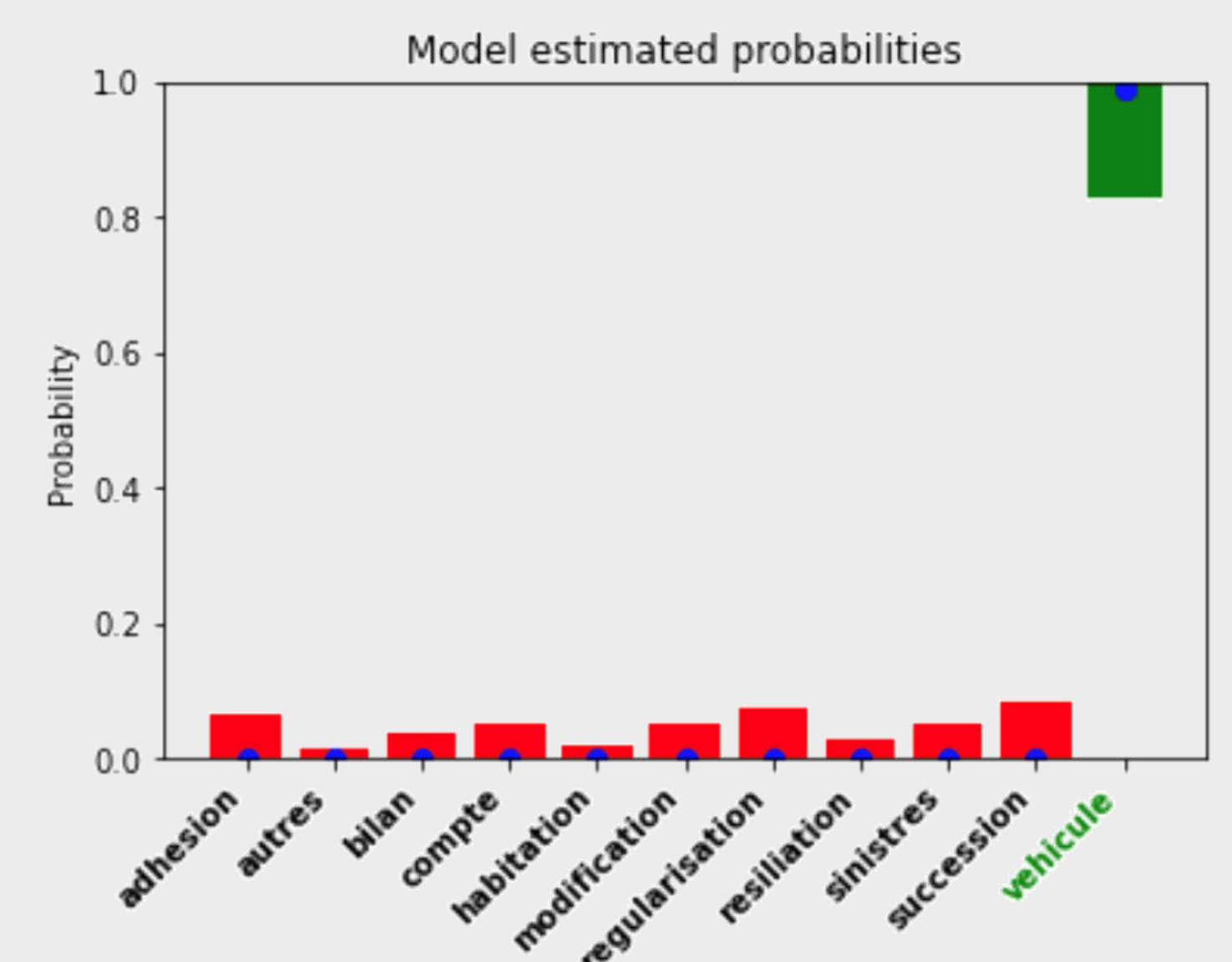
 La librairie TensorFlow Probability permet de construire des réseaux de neurones profonds avec des méthodes probabilistes

Melusine 2.3.6 intègre un nouveau type de modèle avec *Tensorflow Probability* pour estimer une incertitude autour des décisions de routage. Grâce aux intervalles de prédiction, une estimation ponctuelle devient un mode d'un intervalle de prédiction afin de prendre une décision automatique avec une meilleure maîtrise des risques d'erreurs.

Sur 30 simulations sur le movies dataset *IMBD*, le modèle probabiliste atteint la même performance en estimation ponctuelle que le modèle déterministe. Ce modèle probabiliste offre en plus une estimation d'incertitude à chaque prévision.



« Je vous prie de trouver donc la carte grise ainsi que la nouvelle immatriculation. je vous demanderai de faire les changements nécessaires concernant l'assurance. »



5 CONCLUSION

Quantmetry Part of Capgemini Invent a initié plusieurs projets Open Sources dont MAIF pour l'estimation des incertitudes dans les prédictions d'algorithmes de Machine Learning. En participant activement à la nouvelle fonctionnalité sur Melusine permettant d'avoir une estimation incertitude autour de routages d'emails, Quantmetry Part of Capgemini Invent continue d'être acteur d'une IA de confiance et responsable. La participation de Quantmetry dans Melusine s'intègre dans la continuité de son attachement aux projets d'IA valorisables et valorisés par tous, avec l'Open Source comme levier d'innovation.

REFERENCES

<https://github.com/MAIF/melusine>
<https://www.tensorflow.org/probability?hl=fr>