

Stage H/F/X – Speech Analytics et NLP

L'entreprise

Malakoff Humanis est un groupe de protection sociale paritaire, mutualiste et à but non lucratif. Nous sommes ainsi dirigés par des représentants de nos clients. Nous n'avons pas d'actionnaires à rémunérer, l'ensemble de nos bénéfices est donc réinvesti au profit de nos clients, en services, en accompagnement social ou pour soutenir des causes d'intérêt général que nous défendons. Nous sommes un groupe solide financièrement, soucieux d'une gestion rigoureuse et attaché au principe de mutualisation.

Sa raison d'être : « Innover sans cesse au service de l'humain et en faire toujours plus pour protéger et accompagner ses clients entreprises, salariés et retraités. »

Les chiffres clés : N°1 de l'assurance collective (Santé et Prévoyance) en France, 10 millions d'assurés, 426 000 entreprises en assurance collective, 6,3 Milliards de Chiffres d'affaires.

Le Contexte

Le produit Speech Analytics permet de traiter au fil de l'eau les conversations sur nos différents centres d'appel. Une pipeline d'acquisition et de traitement est en place et est adossé au système d'information de l'entreprise (connaissance client, actes de gestions). Cette pipeline alimente différents algo prédictifs (évaluation de la satisfaction, gestion des réclamations, risque de résiliation, etc.).

Le but de ce stage est de développer un nouveau système de classification des conversations qui permettrait d'évaluer la vulnérabilité des entreprises dans notre portefeuille et d'enrichir un produit data déjà existant.

Votre mission

Au sein du département Data Science et IA, vous rejoignez l'équipe produit Speech Analytics deux Data Scientists et d'un Product manager. Votre mission se décompose en trois volets : exploration, analyse et modélisation

Exploration et Analyse des données audios

Le ou la stagiaire devra assurer le maintien de la cohérence des données de conversations audios et participer à l'entreprise de refactorisation des méthodes de traitement de la parole. Vous interviendrez notamment sur l'établissement de lexiques à l'aide de traitement de données de transcriptions et de modèles de reconnaissance de la parole. Vous développerez des métriques de la qualité des données et des métadonnées audios.

Techniques / Technologies : procédés d'analyse de la qualité de la parole modèles de systèmes de reconnaissances simples (types mots clés sur des flux audios), modèles de paroles (type HTK ou Sphinx), modèle de détection de mots clés sur l'audio.

Exploration et Analyse des données temporelles et textuelles

Vous participerez activement au développement de modèles de prédiction (a minima régression linéaire, séries temporelles, ...) pour l'analyse des flux d'appels. Vous interviendrez également sur le traitement des langues à partir du texte pour la génération automatique de résumés (état de l'art, prototypage, implémentation).

Techniques / Technologies : méthodes de machine Learning sur tout type de données (indices numériques, textuelles etc...), des implémentations déjà réalisées dans le cadre des études seront fortement appréciées. Systèmes et processus d'exploitation des données textuelles pour la classification, l'appairage ou le groupement en population (connaissances préalable utiles).

Modélisation et aide à la structuration des données

Vous appuierez le développement de graphes de connaissances (incl. analyse et visualisation de données en graphe pour aider l'émergence de nouvelles dynamiques clients ou interne. Vous pourrez également sur les évolutions de la pipeline (refactorisation, automatisation, monitoring, etc.) ou la MAJ de certains modèles pour améliorer la robustesse, la performance ou la maintenabilité des produits.

Techniques / Technologies : ingénierie des connaissances, méthodes d'analyses de graphes, traitement de données textuelles en masse, bibliothèques de traitement automatique des langues.

Le profil

Nous recherchons une personne capable de développer de nouvelles solutions en s'inspirant de méthodes existantes ou qui imaginera de nouveaux algorithmes adaptés.

Vous préparez un bac+5 (ingénieur, master) en vision par ordinateur, machine learning, traitement du signal ou mathématiques appliquées. Vous avez une connaissance minimale de frameworks du domaine (Scikit-Learn, TensorFlow, PyTorch, Keras, ...).

Vous développez vos propres scripts en autonomie (idéalement Python). Nous utilisons Dataiku pour le prototypage et la mise en production des premières versions des produits. Nous travaillons en mode agile.

Vous êtes créatif et vous souhaitez travailler sur un des défis du moment en machine learning. Vous êtes curieux, vous avez envie d'apprendre. Vous avez du plaisir à utiliser des algos pour tester vos idées. Vous n'avez pas peur d'essayer et vous êtes toujours à la recherche du bon équilibre entre rigueur et exploration. Vous êtes autonome et vous pensez que la partie est plus belle quand on joue avec les autres

Contrat : Stage, 6 mois

Contact : nikola.lackovic@malakoffhumanis.com

Localisation du poste : Paris, 13ème arrondissement – proche BNF et Station F