

Sujet – Industrialisation d'un système de notation automatique de productions écrites en français langue étrangère

Nom du tuteur : Dominique CASANOVA

Entreprise ou organisme : Le français des affaires de la CCI Paris Ile-de-France

1. Contexte

Représentant près de 800 000 entreprises, soit 29% du PIB national, la Chambre de commerce et d'industrie de région Paris Ile-de-France (CCIR) est activement engagée aux côtés des acteurs qui font l'économie régionale – qu'ils soient chefs d'entreprise, décideurs, élèves, apprentis, étudiants ou encore salariés en formation continue.

Sur l'ensemble d'un territoire qui comprend Paris, la Seine-et-Marne, Versailles-Yvelines, l'Essonne, les Hauts-de-Seine, la Seine-Saint-Denis, le Val-de-Marne et le Val d'Oise, la Chambre de région a pour missions de représenter les entreprises pour favoriser leur croissance, de former les hommes et les femmes aux défis de demain, de faire grandir les projets d'entreprise et de promouvoir la région capitale pour accroître son rayonnement.

Créé en 1958, Le français des affaires est un acteur historique et pionnier de la certification et de la formation, en français à visée professionnelle. Son ambition est de promouvoir un français utile et professionnel. Sa mission est de promouvoir le français comme outil des échanges économiques. Le département du développement scientifique du Français des affaires veille au développement de la qualité des tests et certifications au moyen d'analyses statistiques et d'une activité recherche appliquée. Il souhaite exploiter dans les pratiques d'évaluation les possibilités offertes par le traitement automatique des langues et l'apprentissage automatique.

2. Profil recherché :

Élève de 3e année d'école d'ingénieur et étudiant de Master 2 en Statistiques / Mathématiques appliquées, vous avez un intérêt prononcé pour l'apprentissage automatique (machine learning), des compétences éprouvées de programmation en R. Autonome et organisé, vous savez conduire un projet et le finaliser dans les temps. Bon communicant, vous savez présenter la nature de vos travaux à un public non initié et vous intégrer dans une équipe pluridisciplinaire.

3. Mission

Le français des affaires a élaboré un premier prototype pour la prédiction du score et du niveau de copies d'expression écrite d'un test international de français langue étrangère. Votre mission sera d'analyser, d'améliorer et de déployer en production ce prototype en l'intégrant à la chaîne de correction des épreuves.

4. Méthodologie envisagée

Analyse critique du prototype, automatisation de la chaîne de traitement des textes (normalisation des textes, analyse textuelle au moyen de bibliothèques de traitement automatique de la langue, extraction d'indices caractérisant les productions, prédiction du score et du niveau selon le modèle retenu), uniformisation des règles de programmation, généralisation aux différents formats de l'épreuve (pondération et analyse de la pertinence des variables selon le format, apprentissage automatique), évaluation des performances, amélioration de composantes de la chaîne de traitement, intégration à la chaîne de correction. Programmation en R. Base de données : productions écrites d'un test international de français langue étrangère.

5. Résultats attendus

Le programme réalisé générera, à partir de fichiers textes comportant les productions écrites de candidats, une prédiction du score et du niveau de chacun des candidats qui sera intégrée à la base de données psychométrique du test d'évaluation de français. Le rapport décrira le(s) modèle(s)

retenu(s) pour les prédictions, les variables influençant le plus le résultat et rendra compte de la performance du programme en comparaison avec l'accord entre évaluateurs.

6. Bibliographie

Cynthia Gilles (2019). Correction automatique de productions écrites en français langue étrangère. Rapport de stage de 2^e année ENSAI.

Benjamin Larvaron (2019). Correction automatique de productions écrites en français langue étrangère par une approche analytique. Rapport de stage de 2^e année ENSAI.

Informations pratiques :

- Durée du stage : 4 à 6 mois entre le 1^{er} février et le 31 juillet 2019
- Gratification : 3.75 € par heure
- Date limite de candidature : 31 décembre 2019
- Contact : Dominique Casanova (dcasanova@cci-paris-idf.fr)