Projet ML

Parkinson Telemonitoring

Clément Peillon

SIGMA Clermont

Mastère Spécialisé Expert en Science des Données

La maladie de Parkinson touchait en 2019 plus de 8.5 millions de personnes dans le monde selon l’OMS. Cette maladie se manifeste progressivement chez les individus et peut représenter un réel handicap. Le suivi des symptômes est onéreux pour les hôpitaux, nécessite une mobilisation importante de personnel soignant spécialisé et exige la présence du patient pour réaliser des tests souvent chronophages. Pour faire face à ces problèmes, des chercheurs de l’université d’Oxford (Tsanas Athanasios, 2009) ont proposé une méthode de suivi non-invasive basée sur l’analyse de tests de langage et ont mis en ligne une partie de la base de données anonimisée qu’ils ont constituée. Dans ce rapport, on se propose d’étudier différents algorithmes de machine learning pour tenter de prédire l’évolution de la maladie chez les patients.

# Présentation du dataset

Accessible sur le site de [UC Irvine Machine Learning Repository](https://archive.ics.uci.edu/dataset/189/parkinsons+telemonitoring), ce dataset regroupe les données de 42 patients recrutés pour une campagne de test sur six mois afin de mesurer l’évolution de leur maladie de Parkinson. Ces tests présentent deux avantages majeurs. D’abord, ils sont non-invasifs donc le patient n’est pas affecté physiquement et sera moins enclin au stress. De plus, comme il s’agit d’enregistrement de la voix, ces tests peuvent être simplement réalisés à la maison sans avoir besoin de se déplacer dans un hôpital ou une clinique. Cela bénéficie à la fois au patient mais aussi au service hospitalier qui gagnent du temps pour traiter d’autres patients.

Les tests effectués environ tous les deux jours sur la période ont permis de récolter 5875 données. Chaque test consiste en un enregistrement vocal duquel les chercheurs ont extrait 16 caractéristiques détaillées ci-dessous :

* **Jitter (%), Jitter (Abs), Jitter:RAP, Jitter:PPQ5, Jitter:DDP:** Mesures de la variabilité de la fréquence fondamentale (F0) de la voix, qui est souvent altérée chez les patients atteints de la maladie de Parkinson. Un *jitter* élevé indique une instabilité vocale.
* **Shimmer, Shimmer (dB), Shimmer:APQ3, Shimmer:APQ5, Shimmer:APQ11, Shimmer:DDA:** Mesures de la variabilité de l'amplitude de la voix. Un *shimmer* élevé est associé à une voix plus rauque ou instable, un symptôme courant chez les patients parkinsoniens.
* **NHR, HNR**: rapport harmonique-bruit. Mesure de la proportion d'énergie harmonique par rapport au bruit dans le signal vocal. Un HNR plus faible indique une voix plus bruyante et moins harmonieuse.
* **RPDE, DFA, PPE**: mesures non linéaires de la dynamique vocale.

En plus de ca age test\_time et sex puis variables à expliquer

# Annexe