 

**Master Économétrie et Statistiques, parcours Économétrie Appliquée**

**Machine Learning sous Python**

**SVM et réseaux de neurones**

**MEZIANI Sara**

**MORTREUIL Louis**

**Année Universitaire : 2021/2022**

**SOMMAIRE**

1. **PRESENTATION DE L’ETUDE**

Depuis la fin de 2019, le monde ne fonctionne plus comme avant en raison de la pandémie du Covid 19 qui a bouleversé les systèmes économiques, sociaux, éducatifs, etc. L'un des principaux impacts est celui des étudiants du monde entier. Ils ne peuvent pas aller à l'école pour apprendre en face à face, ce qui a un impact important sur leur vie sociale. En effet, selon la Banque Mondiale, près de 1,5 milliard d’élèves dans plus de 170 pays ne vont plus à l’école, leurs établissements ayant été fermés par les gouvernements pour lutter contre la propagation du virus. Dans ce contexte, les ministères de l’éducation du monde entier tentent désormais d’assurer la continuité des apprentissages par le biais de l’enseignement à distance. Dans la plupart des cas, cela implique l’utilisation de plateformes numériques et d’outils technologiques dédiés à l’éducation dans le but de rendre les espaces d’apprentissage aussi ouverts et stimulants que possible.

Aux Etats Unis , la pandémie de COVID-19 a perturbé l'apprentissage de plus de 56 millions d'étudiants.Jusqu'à aujourd'hui, les inquiétudes concernant la fracture numérique exacerbée et la perte d'apprentissage à long terme parmi les apprenants les plus vulnérables d'Amérique continuent de croître.

L’objectif de notre étude est de participer à la compétition Kaggle “ LearnPlateform Covid-19 Impact On Digital Learning “ pour analyser et déterminer l’impact de la Covid 19 sur l’apprentissage des étudiants tout en utilisant les données numériques aux USA et présenter notre démarche et nos résultats obtenus.

Pour répondre à cette problématique, nous allons utiliser dans ce travail plusieurs méthodes de machine learning. Les données que nous allons exploiter trois fichiers de données fournis par les organisateurs de la compétition et qui sont :

* **Les données engagement**: sont basées sur l'extension Student Chrome de LearnPlatform. L'extension collecte les événements de chargement de page de plus de 10 000 produits de technologie éducative dans notre bibliothèque de produits, notamment des sites Web, des applications, des applications Web, des logiciels, des extensions, des livres électroniques, des matériels et des services utilisés dans les établissements d'enseignement. Les données d'engagement ont été agrégées au niveau du district scolaire, et chaque fichier représente les données d'un district scolaire.
* **Les données products**: comprend des informations sur les caractéristiques des 372 meilleurs produits avec la plupart des utilisateurs en 2020.
* **Les données districts**: comprend des informations sur les caractéristiques des districts scolaires, y compris des données du NCES et de la FCC.

**L’Etat de l’apprentissage numérique en 2020 (Analyse exploratoire)**