|  |
| --- |
| **Projets PFA** |
|  |

# CHRONOLOGIE DU DOCUMENT :

|  |  |  |  |
| --- | --- | --- | --- |
| **Etudiant** | **Action** | **Date** | **Version** |
| Asmae GHALAS  Mouad ELKACEM | Création de document | 12/05/2020 | V1 |

# INTITULE DU PROJET :

|  |  |  |  |
| --- | --- | --- | --- |
| **Titre du projet :** | | | |
| **Encadrant** | **Co-Encadrant** | **Etudiant** | **Date de début** |
| Dr EL ASRI Bouchra |  | GHALAS Asmae  ELKACEM Mouad |  |

# DESCRIPTION DU PROJET :

# La désinformation représente une menace sérieuse pour la société, car la prolifération de fausses informations peut avoir un impact significatif sur la perception de la réalité par un individu. Les fausses nouvelles sont une affirmation ou une histoire fabriquée, avec l'intention de tromper, souvent pour un objectif secondaire tel qu'un gain économique ou politique. À l'ère des nouvelles numériques et des médias sociaux, les fausses nouvelles peuvent se propager rapidement, affectant de grandes quantités de personnes en peu de temps. Pour atténuer l'impact négatif des fausses nouvelles sur la société, diverses organisations emploient désormais du personnel pour vérifier les allégations douteuses grâce à une procédure manuelle de vérification des faits, cependant, ce processus est très laborieux.

Dans ce projet intitulé ‘’ Développement d’un système d’acquisition et de vérification de données dans un système de veille’’, on va proposer une solution de détection des fausses nouvelles, cette solution est basée sur le Machine Learning, l'une des tendances technologiques.

# L'ARCHITECTURE FONCTIONNELLE/TECHNIQUE :

# 

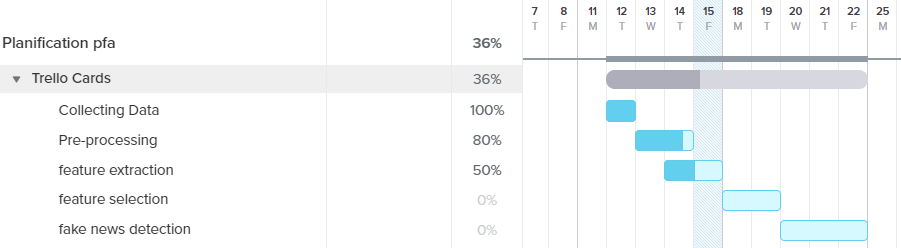
Ce projet va permettre par la suite aux différents utilisateurs de vérifier la véracité d’une information donnée par ce dernier à l’aide d’une interface. La vérification de l’information va être effectuer à l’aide d’un model construit à partir d’un algorithme choisi et qui nous assures un niveau de précision élevé.

Pour effectuer un tel projet, on propose le plan suivant :

* La collection des données :  vise à rassembler des articles pertinents à partir de diverses sources (Kaggle, Google datasets …).
* Le prétraitement des données : obtenir les données dans un format standard propre pour une analyse plus approfondie. Par exemple supprimer la ponctuation, rendre les lettres minuscules, supprimer les chiffres
  + Outils : Tokenization, Lemmatization …
* Feature extraction : vise à réduire le nombre d'entités dans un ensemble de données en créant de nouvelles entités à partir des entités existantes (puis en supprimant les entités originales). Ces nouveaux ensembles de fonctionnalités réduits devraient alors être en mesure de résumer la plupart des informations contenues dans l'ensemble de fonctionnalités d'origine. De cette façon, une version résumée des fonctions originales peut être créée à partir d'une combinaison de l'ensemble d'origine.
  + Outils : Bag-of-Words, TF-IDF, N-Gramms …
* Feature selection : C’est le processus de recherche et de sélection des fonctionnalités les plus utiles dans un ensemble de données, c’est une étape cruciale du pipeline d'apprentissage automatique. Les fonctionnalités inutiles réduisent la vitesse d'entraînement, diminuent l'interprétabilité du modèle et, surtout, diminuent les performances de généralisation sur l'ensemble de test.
  + Outils : Univariate selection, Recursive Feature Elimination …
* L’entrainement du « Predictive Model » :  vise à utiliser les informations de position (la perception de chaque journal face à un sujet) pour déterminer la véracité de la réclamation.
  + Outils : Random Forest, K Nearest Neighbors …

# DIAGRAMME DE GANTT :

# 



# PLANIFICATION DES ACTIVITES :

[Police : Garamond, Style : Normal, Taille :12, Colleur : Noir]

|  |  |  |
| --- | --- | --- |
| **Activité** | **Livrable(s)** | **Date prévue de livraison** |
|  |  | XX/XX/2020 |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

# AVANCEMENT DES TRAVAUX :

[Police : Garamond, Style : Normal, Taille :12, Colleur : Noir]

|  |  |  |  |
| --- | --- | --- | --- |
| **Semaine** | **Activités Réalisées/En cours** | **Avancement** | **Estimation en J/H** |
| **XX au XX/XX/2020** |  | % |  |
|  | % |  |
|  | % |  |
|  | % |  |
|  | % |  |
|  | % |  |

# LIVRABLES REALISES :

[Police : Garamond, Style : Normal, Taille :12, Colleur : Noir]

|  |  |
| --- | --- |
| **Livrable I** |  |
| **Livrable II** |  |
| **Livrable N** |  |