



République Algérienne Démocratique et Populaire



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université AMO de Bouira

Faculté des Sciences et des Sciences Appliquées

Département d'Informatique

# Mémoire de Master

## en Informatique

*Spécialité : Génie des Systèmes Informatique*

## Thème

---

Evaluation automatique de réponses courtes en  
langue Arabe par l'apprentissage automatique

---

Encadré par

— PR.DJAMEL BENNOUAR

Réalisé par

— LARIBI OUEDAD

— LADJAL SARAH

2020/2021



Département d'informatique

قسم الاعلام الآلي

## AUTORISATION DE SOUTENANCE DE MASTÈRE

Monsieur le Chef du département Informatique ;

Je vous prie de bien vouloir accepter le dépôt du dossier de soutenance des l'étudiants(es) :

Matricule	Nom	Prénom
	Lodjal	Sarah
	Laribi	Ouedad

qui ont, sous mon encadrement, réalisé le sujet intitulé :

**Evaluation automatique de réponses courtes en Langue Arabe par l'apprentissage automatique**

Je propose comme jury de soutenance :

- Président : .....
- Examineur : .....
- Encadreur : ...*Djamal BENNOUAR*.....
- Invité : .....

L'encadreur :

Djamal BENNOUAR

## تصريح شرفي خاص بالالتزام بقواعد النزاهة العلمية لإنجاز بحث<sup>1</sup>

لحجال صدارة	أنا المعني أسفل السيد(ة) :
طالب	الصفة : طالب أستاذ باحث
119970309005540009	الحامل لبطاقة التعريف الوطنية رقم :
2016.04.09	الصادرة بتاريخ:
المسجل(ة) بكلية العلوم والعلوم التطبيقية، قسم الإعلام الآلي، والمكلف(ة) بإنجاز أعمال بحث (مذكرة تخرج، مذكرة ليسانس، مذكرة ماجستير، أطروحة دكتوراه)، عنوانها : Evaluation automatique des réponses courtes en langue Arabe par l'apprentissage automatique	

أصرح بشرفي أنني أتزم بمراعات المعايير العلمية والمنهجية ومعايير الأخلاقيات المهنية والنزاهة الأكاديمية المطلوبة في إنجاز البحث المذكور أعلاه.

التاريخ 10-09-2021

توقيع المعني(ة)



## تصريح شرفي خاص بالالتزام بقواعد النزاهة العلمية لإنجاز بحث<sup>2</sup>

أنا المعني أسفله السيد(ة) : Laribi Ouedad

الصفة : طالب أستاذ باحث Etudiant

الحامل لبطاقة التعريف الوطنية رقم : 5258-562655

المصادرة بتاريخ: 2015.04.29

المسجل(ة) بكلية العلوم والعلوم التطبيقية، قسم الإعلام الآلي،  
والمكلف(ة) بإنجاز أعمال بحث (مذكرة تخرج، مذكرة ليسانس، مذكرة ماجستير، أطروحة دكتوراه)، عنانها :

Master

أصبح بشرفي أني ألتزم بمراعات المعايير العلمية والمنهجية ومعايير الأخلاقيات المهنية والنزاهة الأكاديمية المطلوبة في إنجاز البحث المذكور أعلاه.

التاريخ 10-09-2021

توقيع المعني(ة)



## Remerciements

2 حسب القرار رقم 1082 المؤرخ في 27 ديسمبر 2020 والذي يحدد القواعد المتعلقة بدولة من الشراكة العلمية وبتفاتها

# *Remerciements*

Nous tenons tout d'abord à remercier Allah, qui nous a donné la force, la capacité et la patience pour accomplir ce travail.

Nous tenons a remerciements avec un grand plaisir et un grand respect a notre encadreur Pr BENNOUAR Djamel, pour ses conseils, sa disponibilité et ses encouragements qui nous ont permis de réaliser ce travail dans les meilleures conditions.

Nous tenons a remercier les membres du jury pour avoir fait le plaisir d'accepter d'examiner ce travail. Nous remercions nos amies pour leur soutien et encouragement.

Nous souhaitons adresser nos remerciements les plus sincères au corps professoral et administratif de l'Université Akli Mohand Oulhadj, Bouira, pour la richesse et la qualité de leur enseignement et qui déploient de grands efforts pour assurer à leurs étudiants une formation actualisée.

Nous tenons à remercie toute personne qui a participé de près ou de loin à l'exécution de ce modeste travail.

# *Dédicaces*

J'ai le plaisir de dédier ce travail à mes très chers parents à qui est l'expression de mes profonds sentiments de respect, de gratitude et de reconnaissance que dieu les garde pour nous. A mes frères et mes sœurs, A mon neveu Bilal, A tous la famille. A tous mes amis. A tous ceux qui m'aiment et que j'aime.

*Ladjal sarah*

## *Dédicaces*

Je dèdie ce travail Au meilleur des pères merci beaucoup pour tout le soutien et les sacrifices que vous avez consentis pour moi ,A mes frères et ma sœur,A toute ma famille ,A tous mes collègues et amies , A tous ceux qui m'aiment et que j'aime.

*Laribi ouedad.*

# Table des matières

<b>Table des matières</b>	<b>i</b>
<b>Table des figures</b>	<b>iv</b>
<b>Liste des tableaux</b>	<b>vi</b>
<b>Liste des abréviations</b>	<b>vii</b>
<b>1 introduction générale</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.2 Problématique . . . . .	1
1.3 Objectif du travail . . . . .	2
1.4 Importance de notre travail . . . . .	2
1.5 Structure du mémoire . . . . .	3
<b>2 Etat de l’art</b>	<b>4</b>
2.1 Introduction . . . . .	4
2.2 Les systèmes ASAG : . . . . .	4
2.2.1 Généralités sur les ASAG : . . . . .	4
2.2.2 Une vue historique sur les systèmes ASAG : . . . . .	5
2.3 Les approches de similarité des ASAG . . . . .	9
2.3.1 Similarité syntaxique : . . . . .	9
2.3.2 Similarité sémantique : . . . . .	12
2.3.3 Similarité hybride : . . . . .	13
2.4 Le Word embedding : . . . . .	13

2.4.1	Continuous Bag Of Word(CBOW) : . . . . .	13
2.4.2	Skip-Gram(SG) : . . . . .	14
2.5	Concepts fondamentaux liées à l'apprentissage automatique : . . . . .	15
2.5.1	Définitions : . . . . .	15
2.5.2	Types de l'apprentissage automatique : . . . . .	16
2.6	Métriques d'évaluation des modèles : . . . . .	20
2.7	Les Revue de la littérature . . . . .	25
2.7.1	Systèmes qui utilisent le Machine Learning . . . . .	25
2.7.2	Systèmes pour la langue Arabe . . . . .	25
2.8	Conclusion . . . . .	26
<b>3</b>	<b>Système d'évaluation automatique des réponses courtes</b>	<b>27</b>
3.1	Introduction . . . . .	27
3.2	Méthodologie . . . . .	27
3.3	Les modèles proposés . . . . .	28
3.4	Mise en œuvre des modelés . . . . .	29
3.4.1	Prétraitement de Données : . . . . .	29
3.5	Les Etapes d'élaboration d'espace sémantique : . . . . .	32
3.5.1	Collection des informations spécifiques de domaine : . . . . .	32
3.5.2	Prétraitement de corpus : . . . . .	33
3.5.3	Génération des termes uniques : . . . . .	33
3.5.4	Elaboration de Matrice d'espace sémantique : . . . . .	33
3.6	L'approche Machine Learning : . . . . .	34
3.6.1	Sélection des caractéristiques : . . . . .	35
3.6.2	Partitionnement des données : . . . . .	35
3.6.3	Entraînement du modèle : . . . . .	36
3.6.4	Test du modèle : . . . . .	36
3.6.5	Evaluation : . . . . .	36
3.7	Conclusion . . . . .	36
<b>4</b>	<b>Résultats expérimentaux et évaluation</b>	<b>37</b>
4.1	Introduction . . . . .	37
4.2	Implémentation : . . . . .	37

4.2.1	DataSet : . . . . .	37
4.2.2	Structuration de DataSet : . . . . .	38
4.2.3	Les techniques utilisées dans le prétraitement : . . . . .	40
4.2.4	Extraction des caractéristiques : . . . . .	42
4.2.5	Processus de l'apprentissage automatique : . . . . .	46
4.2.6	Environnement du développement : . . . . .	47
4.3	Résultat et discussion : . . . . .	47
	<b>Conclusion générale</b>	<b>49</b>
	<b>Bibliographie</b>	<b>50</b>

# Table des figures

2.1	Organisation historique des systèmes ASAG [3]. . . . .	5
2.2	Les approches de similarité. . . . .	9
2.3	Mesures de similarité syntaxique. . . . .	11
2.4	Mesures de similarité sémantique. . . . .	12
2.5	L'architecture de modèle CBOW. . . . .	14
2.6	L'architecture de modèle SkipGram. . . . .	15
2.7	les types d'apprentissage automatique avec leurs algorithmes. . . . .	16
2.8	Phase d'apprentissage du modèle d'apprentissage supervisé. . . . .	17
2.9	Les différents problèmes de l'apprentissage automatique supervisé. . . . .	18
2.10	Modèle général d'un réseau de neurone artificiel . . . . .	20
2.11	Tableau 1 :ASAG Systèmes qui utilisent le Machine Learning . . . . .	25
2.12	Tableau 2 :ASAG Systèmes qui utilisent le Machine Learning la langue arabe . . . . .	26
3.1	Plan suivi dans la conception . . . . .	28
3.2	Représentation des Modèles proposés . . . . .	28
3.3	Représentation des différentes étapes de prétraitement des données . . . . .	29
3.4	Matrice d'espace sémantique . . . . .	33
3.5	Mécanisme suivi dans l'apprentissage automatique de notre système . . . . .	34
3.6	Processus de l'extraction des caractéristiques . . . . .	35
4.1	Tableau 4.1 dataset . . . . .	38
4.2	fichier XML qui représente l'ensemble de données« AR-ASAG Dataset. . .	39
4.3	fichier XML qui représente « Questions-Réponses Modèle AR-ASAG Dataset. .	40

4.4	Exemple de normalisation . . . . .	41
4.5	Exemple de StopWord . . . . .	41
4.6	Architectures de modèles CBOW et skip-gram . . . . .	44
4.7	Tableau 4.2 : Représentation vectorielle des mots . . . . .	45
4.8	Tableau 4.3 : resultat . . . . .	47

# Liste des tableaux

# Liste des abréviations

ASAG Système d'évaluation automatique des réponses courtes(Automatic Short Answer Grading)

ANN Réseau neurone artificiel(Artificial neural network)

CAH Clustering ascendant hiérarchique

CBOW CBOW Sac continu de mots (Continuous Bag Of Words)

CP Calcul de Pearson

DTR Decision tree regression

DT Arbre de décision (Decision tree)

ES Espace Sémantique

LR Régression Linéaire (Linear Regression)

ML Apprentissage automatique (Machine learning)

WE Word Embedding

RMSE Erreur quadratique moyenne (Root Mean Squared Error)

SG Skip-Gram

SAMText Short Answer Measurement of TEXT

SVR Prise en charge de la régression vectorielle( support vector regression)

SVM Machine à vecteur de support (Support Vector Machine)

VSM Vector Space Model

## introduction générale

### 1.1 Introduction

Aujourd'hui, les systèmes de télé-enseignement semblent avoir atteint un bon degré de maturité qui permettrait de drainer un nombre de plus en plus important d'enseignants à l'utilisation de ces techniques pour d'une part la diffusion des ressources pédagogique, la communication plus efficace entre apprenant, enseignant et concepteur de cours et d'autre part le contrôle continu des étudiants.

Le contrôle continu à travers les tests en ligne semble être le point le plus attractif pour les enseignants. Cependant ce point ne semble pas avoir atteint un haut degré de maturité pour qu'il puisse devenir un facteur important d'attraction des enseignants à l'utilisation des plateformes de télé-enseignement.

Le travail est orienté vers la langue arabe qui bien que largement utilisée aujourd'hui, n'a pas encore bénéficié de recherche et de résultats matures dans le domaine de l'évaluation automatique. Notre système final sera testé avec la langue anglaise afin d'obtenir un système complet, fiable et standard.

### 1.2 Problématique

L'idée de base de l'évaluation automatique consiste à l'étudiant est amené à écrire un à trois phrases (réponses courts), semble être une technique qui permettrait d'évaluer

de manière assez efficace l'acquisition des connaissances d'un cours, et permettre à l'enseignant de voir quels sont les points acquis et les points de difficulté pour les étudiants.

Actuellement, les outils d'évaluation automatique de réponses courtes ne semblent pas avoir atteint un haut degré de maturité et impose aux examinés des compétences et des contraintes importantes pour leur exploitation. En plus du fait qu'ils soient orientés pour la plupart vers une langue particulière (Anglais) et vers des matières particulières, les outils actuels ne sont pas efficaces dans un environnement réel qui est souvent hostile à cause de plusieurs facteurs tels que les erreurs d'orthographe, les erreurs d'inattention (répétition erronée de lettre, l'oubli d'un accent, l'utilisation d'un mauvais déterminant etc).

A l'Ere du Machine Learning nous voulons explorer ces modèles pour l'évaluation automatique en langue arabe et analyser l'impact sur les outils déjà développés dans le contexte de travaux précédents.

### 1.3 Objectif du travail

L'objectif de notre travail consiste à :

- Etudier quelques approches de l'apprentissage automatique d'évaluation automatique de réponses courtes (de quelques mots à trois phrases).
- Maitriser une (ou deux) approche(s) intelligente(s) facile à mettre en œuvre et à programmer.
- Evaluer dans le contexte de la langue arabe la (ou les) approche(s) choisie(s).

### 1.4 Importance de notre travail

Les objectifs de l'utilisation de l'évaluation automatique dans l'éducation comprennent des avantages suivants :

- Fournir aux étudiants des informations détaillées sur leur période d'études d'une manière plus efficace que les évaluations traditionnelles.
- Encourager l'intégration de systèmes de test spécifiques liés à la réponse ouverte dans la plate-forme e-Learning.

- S'adapter à la culture des tests en ligne, qui est un nouvel avantage dans le domaine de l'éducation.
- Faciliter les tâches d'évaluation des enseignants et réduire le temps de correction.
- Un système d'évaluation non supervisé et entièrement automatisé réduire le risque d'erreur humaine et les soupçons de tricherie ou de subjectivité.
- Permet à un chercheur avantage d'exportation, pour construire une banque de question pour l'enseignement.

## 1.5 Structure du mémoire

Notre mémoire est organisé de la manière suivante :

- **Chapitre 1** : Introduction générale qui englobe le contexte et problématique du travail.
- **Chapitre 2** : Etat de l'art qui englobe les concepts de base liés à la notation automatique, ainsi qu'une synthèse des travaux déjà réalisés dans ce contexte.
- **Chapitre 3** : Conception du système de l'Evaluation Automatique des Réponses Courtes la méthodologie utilisée pour mettre en œuvre le système.
- **Chapitre 4** : Présente le cadre pratique de nos travaux, présentation de la mise en œuvre et discussions des résultats obtenus.

Nous terminons notre étude par une conclusion générale et des perspectivesProjetées.

## Etat de l'art

### 2.1 Introduction

Le domaine de la notation automatique comprend de nombreux concepts qui sont considérés comme un pilier pour la construction d'un système performant. Dans ce chapitre nous allons situer notre travail par rapport aux fonctionnements des ASAG, les différentes approches d'évaluation automatique de réponses courtes, ainsi que les revues par rapport à notre travail.

### 2.2 Les systèmes ASAG :

#### 2.2.1 Généralités sur les ASAG :

Les systèmes de notation automatique à réponse courte (ASAG) sont conçus pour évaluer automatiquement les réponses courtes en langage naturel, comportant une longueur de quelques mots à quelques phrases. Dans le monde réel, de nombreux facteurs, tels que le niveau de difficulté et la diversité des réponses des élèves, varient considérablement d'une question à l'autre, ce qui permet à différentes techniques ASAG de devenir supérieures pour différentes mesures d'évaluation.

Les réponses textuelles peuvent être des réponses courtes ou des essais, et chacun doit être évalué différemment. Normalement, les systèmes de notation automatique des réponses courtes évaluent et marquent la réponse d'un apprenant en le comparant à une ou plusieurs réponses correctes. [1]

Les ASAG se basent sur le contenu plutôt que sur le style. Une mauvaise qualité d'écriture

(de formulation) est, jusqu'à un certain point, tolérée facilitant ainsi la production d'une idée, une méthode, ou un produit original qui est la meilleure façon de vérifier l'acquisition des informations et des connaissances selon la taxonomie de Bloom [2] il existe plusieurs taxonomies représentant ces niveaux d'acquisition de connaissances. Dans ce domaine, la taxonomie la plus fréquemment citée est celle de Bloom.

## 2.2.2 Une vue historique sur les systèmes ASAG :

Bien qu'il existe de nombreux systèmes ASAG qui ont été exploités dans le temps par le système pédagogique élaboré, l'un des systèmes les plus récents et les plus complets émerge des « The Eras and Trends of Automatic Short Answer Grading » [3] où les auteurs rassemblent le système automatique. Les systèmes de notation pour les réponses courtes à 5 époques différentes sont illustrés à la Figure 1 avec les outils ASAGS les plus populaires à chaque époque :

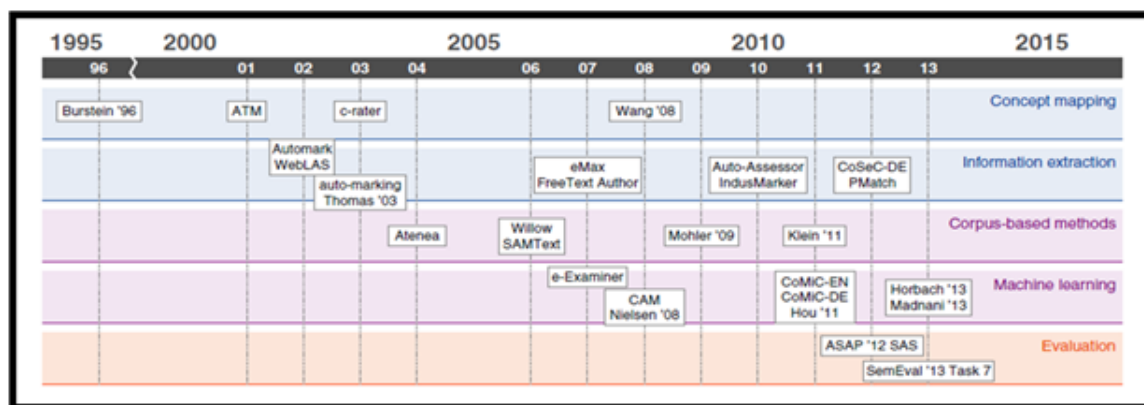


FIGURE 2.1 – Organisation historique des systèmes ASAG [3].

### 2.2.2.1 Les méthodes basées sur le mappage de concepts (Concept Mapping) :

La technique de mappage de concept est apparue en 1996. Cette technique s'exprime au niveau de la phrase. L'idée est de regarder les réponses des étudiants composées de plusieurs concepts et de découvrir la présence ou l'absence de chaque concept lors de l'enregistrement. Nous prenons des exemples de systèmes ASAG :

**Burstein** Burstein et al. (1996) utilisent la technique de la représentation de la structure conceptuelle lexicale selon laquelle un lexique conceptuel et une grammaire conceptuelle doivent être développés à partir d'un ensemble d'apprentissage avant de classer les hypothèses dans les réponses des étudiants.[4]

**C-rater** c'est un moteur de notation automatisé qui a été développé pour marquer les réponses aux questions de réponse courte, il est basé sur le contenu. Il ne s'agit pas simplement d'un programme d'appariement des chaînes, mais plutôt d'un ensemble des règles et une représentation canonique des textes en utilisant la variation syntaxique, anaphore, morphologique variation, synonymes et correction orthographique pour attribuer un score complet ou partiel à une question de réponse courte. C-rater a été utilisé dans deux études : National Assessment for Educational Progress (NAEP) et une évaluation à l'échelle de l'État dans l'Indiana. Dans les deux études, c-rater était d'accord avec les personnes humaines environ 84 % du temps.[5]

### **Les Méthodes basées sur l'extraction d'information (Information Extraction)**

Il s'agit d'une série d'opérations de correspondance de modèles telles que les expressions régulières ou l'analyse d'arborescence. Ils peuvent extraire des données structurées à partir de sources non structurées, telles que du texte libre et représenter les données structurées sous forme de groupes pouvant être utilisés dans de nombreuses applications. Nous prenons des exemples de systèmes ASAG :

**eMax** Le système eMax est conçu pour prendre en charge divers scénarios d'examen, il fournit une évaluation semi-automatique des textes courts librement formulés ainsi que des problèmes mathématiques partiellement résolus. L'approche de la notation est une approche combinatoire, où toutes les formulations possibles sont prises en compte lors de l'appariement de modèles.[6]

**FreeText Author** Le système a été utilisé pour écrire et marquer les tâches d'évaluation de texte libre à réponse courte, il fournit une interface utilisateur graphique pour l'entrée de réponse de l'enseignant et le classement des réponses des élèves. Les réponses de l'enseignant sont composées comme des modèles syntaxiques-sémantiques pour que les réponses de l'élève soient comparées. Ces modèles sont automatiquement générés à partir de l'entrée de langage naturel des réponses des enseignants.[7]

### Les méthodes basées-corpus (Corpus based Methods)

En exploitant les propriétés statistiques d'un grand corpus de documents. Bien que ces méthodes soient souvent utilisées pour des applications avec de longs textes. Ces méthodes peuvent également être utiles lors de l'interprétation de synonymes dans des réponses courtes, en utilisant uniquement le vocabulaire de réponse original et en limitant les réponses correctes qui peuvent être identifiées. Une approche typique pour augmenter le vocabulaire consiste à utiliser des corpus parallèles bilingues pour analyser combien de fois des paires de termes sont résolues dans une seule traduction en langue seconde. Ensuite, des synonymes avec des traductions courantes peuvent être incorporés dans la réponse de l'enseignant. Nous prenons les systèmes ASAG comme exemples :

**SAMText (Short Answer Measurement of TEXT)** applique une variante de LSA basée sur une structure de données inversées d'index, qui est estimée par le contenu d'un analyseur Web utilisant des documents topiques pertinents. En revanche, LSA utilise normalement une structure de données matricielle basée sur de grands corpus pour modéliser la parenté sémantique. L'index inversé et l'idée rampante est plus approprié pour des réponses courtes comparées aux réponses longues parce que les explorations de web peuvent être adaptées à chaque sujet au lieu d'essayer de modéliser toute la langue à la fois.[8]

**Mohler** utilise des techniques non supervisées pour la tâche de classement automatique de réponse courte, en comparant un certain nombre de mesures fondées sur la connaissance et le corpus de la similitude de texte et évaluant l'effet du domaine et de la taille sur les mesures basées sur le corpus, et introduisant également une technique pour améliorer les performances du système en intégrant commentaires des réponses de l'élève.[9]

## Les méthodes basées sur l'apprentissage automatique (Machine Learning)

Une nouvelle technologie est devenue populaire en 2008, la technologie d'apprentissage automatique, ils utilisent généralement un certain nombre de métriques extraites de techniques similaires de traitement du langage naturel, qui sont ensuite combinées en un seul score ou score à l'aide d'un modèle de classification ou de régression. Cela peut être pris en charge par un ensemble d'outils d'apprentissage automatique tel que [10]. Les fonctionnalités impliquant un sac de mots et des n-grammes sont typiques de cette classe, tout comme les arbres de décision et les machines vectorielles prenant en charge les algorithmes d'apprentissage représentatifs. Nous prenons des exemples de systèmes ASAG, les systèmes les plus représentatifs sont :

**CAM** Utilise une approche pour l'évaluation du contenu des réponses des apprenants de langue anglaise aux questions de compréhension de la lecture à réponse courte, il utilise un classificateur k-plus proche voisin (KNN) et des caractéristiques qui mesurent le pourcentage de chevauchement du contenu sur différents niveaux linguistiques entre les réponses de l'enseignant et celles de l'élève. Il atteint une précision de 87% pour la détection d'erreurs sémantiques qui est amélioré jusqu'à 88% sur un ensemble de données de test retenu.[11]

**Madnani 13** il mesure les compétences de compréhension de la lecture des élèves par le marquage automatique des résumés rédigés par les élèves et compare les résultats par rapport à la notation humaine. L'approche d'apprentissage automatique comprend huit caractéristiques (BLEU, ROUGE, mesures concernant différentes dimensions de la copie de texte, nombre de phrases, et nombres de mots de connecteur de discours couramment utilisés) comme entrée à un classificateur de régression logistique.[12]

## 2.3 Les approches de similarité des ASAG

L'évaluation des similitudes entre entités textuelles est l'une des problématiques centrales dans de nombreuses disciplines telles que l'évaluation automatique des réponses, l'analyse de données textuelles, la recherche documentaire, ou encore l'extraction de connaissances à partir de données textuelles (text mining)... Dans chacun de ces domaines, les similarités sont déjà utilisées Pour une variété de traitements. Nous citons trois grandes classes d'approches de similarité (Figure 2) :

1. Similarité syntaxique (String-basedsimilarity).
2. Similarité sémantique (Semanticsimilarity).
3. Similarité hybride (hybridsimilarity)

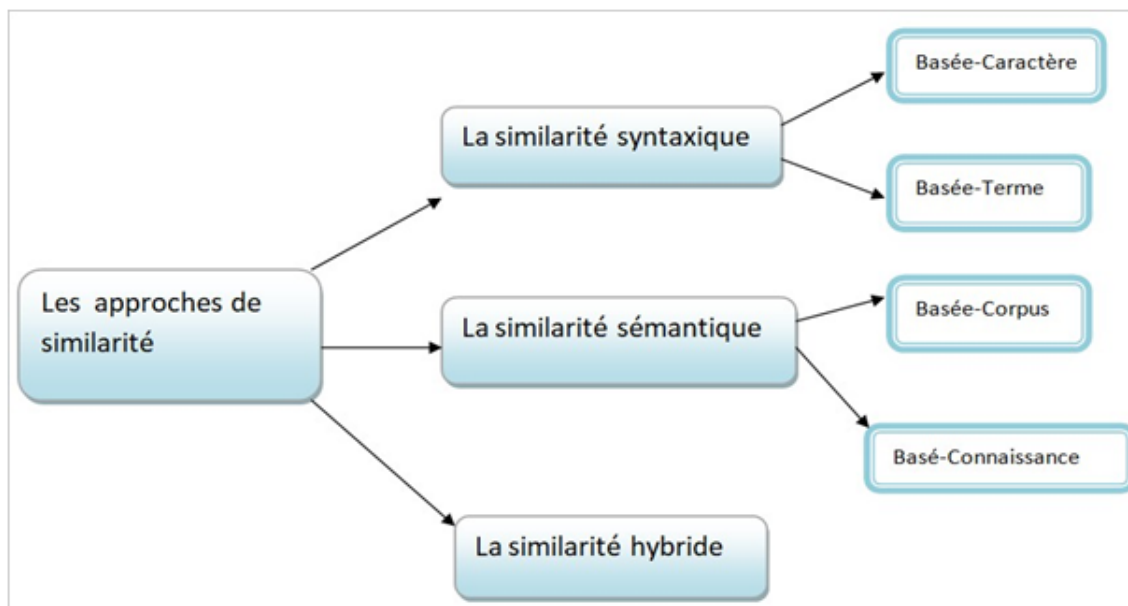


FIGURE 2.2 – Les approches de similarité.

### 2.3.1 Similarité syntaxique :

Elle consiste à comparer les documents textuels, c'est une métrique qui mesure la similarité entre 2 chaînes de caractères. Par exemple, les chaînes de caractère "" et "" peuvent être considérées comme très proches, alors que "" et "" pourront être considérées comme très différents [13]. La figure 3 montre les mesures de similarité syntaxique. Parmi ses mesures, sept sont basées sur des caractères tandis que les autres sont des mesures de

distance basées sur des termes. Parmi ces mesures nous présentons :

**Indice de Jaccard :** Le rapport entre la cardinalité de l'intersection des ensembles considérés et la cardinalité de l'union des ensembles. Il permet d'évaluer la similarité entre les ensembles. Les documents d1 et d2 sont donc représentés non pas comme des vecteurs, mais comme des ensembles de termes. La similarité obtenue (d1, d2) [0, 1]. [14]

$$J(A, B) = \frac{(A \cap B)}{(A \cup B)}$$

**Distance euclidienne :** La distance euclidienne calcule la similarité entre deux documents d1 et d2 comme la distance entre leurs représentations vectorielles ramenées à un seul point [15]. Où n est le nombre total de termes représentés, i.e. la taille des vecteurs.

$$Sim_{euclidienne} = \sqrt{\sum_{i=1}^n (d1_i - d2_i)^2}$$

**Cosinus :** La similarité cosinus est fréquemment utilisée [16] en tant que mesure de ressemblance entre deux texte d1 et d2. Il s'agit de calculer le cosinus de l'angle entre les représentations vectorielles des textes à comparer. La similarité obtenue (d1, d2) [0, 1].

$$Sim_{cosinus} = \frac{\vec{d1} \cdot \vec{d2}}{\|\vec{d1}\| \|\vec{d2}\|}$$

**Indice de Dice :** L'indice de Dice mesure la similarité entre deux documents d1 et d2 en se basant sur le nombre de termes communs à d1 et d2 [13]. Où Nc est le nombre de

$$sim_{dice} = \frac{2 Nc}{N1 + N2}$$

termes communs à d1 et d2, et N1 (resp. N2) est le nombre de termes de d1 (resp.d2).

**Coefficient de corrélation de Pearson :** Le coefficient de corrélation de Pearson calcule la similarité entre deux documents d1 et d2 comme le cosinus de l'angle entre leurs représentations vectorielles centrées- réduites. La similarité obtenue (d1, d2) [1, 1][13].

$$sim_{pearson}(d1, d2) = sim_{cosinus}(d1 - \bar{d1}, d2 - \bar{d2})$$

D'où (resp. ) représente la moyenne de  $d1$  (resp.  $d2$ ).

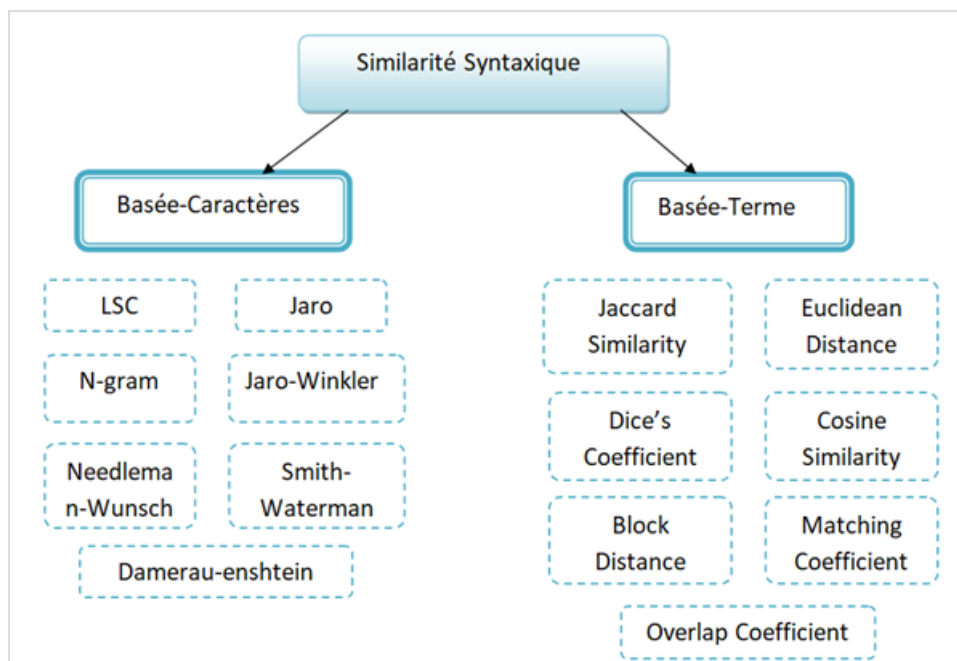


FIGURE 2.3 – Mesures de similarité syntaxique.

Afin de mesurer la similarité syntaxique, il est nécessaire de construire un modèle spatial vectoriel (vector space model). Il réduit la complexité des documents et facilite leur manipulation c'est-à-dire la version en texte intégral en un vecteur qui décrit le contenu du document. Le document est représenté sous forme vectorielle en deux étapes :

**Extraction des termes pertinents :** Cela inclut le prétraitement du texte des documents texte en supprimant les mots vides, la ponctuation et tout « retour de ligne », de lemmatiser le texte et de le segmenter.

**Calcul des poids :** Le poids de chaque terme dans un document peut être obtenu de différentes manières : booléenne, fréquence des termes, Tf-Idf (Term frequency – Inverse Document Frequency). [17]

### 2.3.2 Similarité sémantique :

La similarité sémantique est un concept selon lequel un ensemble de documents ou de termes se voient attribuer une métrique basée sur la ressemblance de leur signification (contenu sémantique) [13]. La figure 4 montre les mesures de similarité sémantique les plus répandues.

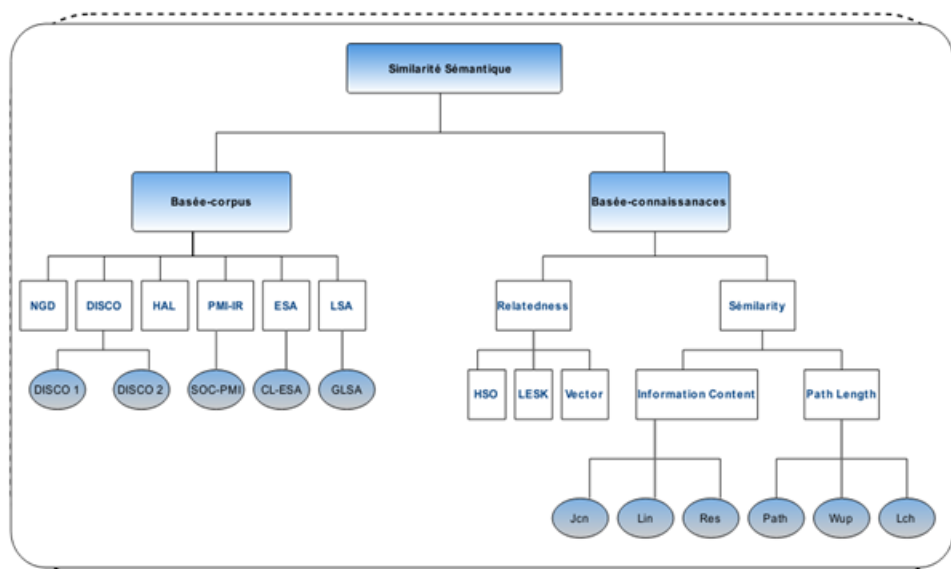


FIGURE 2.4 – Mesures de similarité sémantique.

D'après la figure 4 nous distinguons deux types de similarité sémantique :

#### Les mesures basées sur des corpus :

C'est une mesure statistique qui détermine la similarité entre les mots en fonction de l'information obtenue d'un corpus volumineux. Elle consiste à créer un espace sémantique à partir des cooccurrences de mots. Une matrice mot à mot est formée dont chaque élément de la matrice est la force d'association entre le mot représenté par la ligne et le mot représenté par la colonne. Les valeurs matricielles sont accumulées en pondérant la cooccurrence de manière inversement proportionnelle à la distance de focalisation du mot. Les mots voisins les plus proches sont considérés comme reflétant davantage la sémantique du mot cible et sont donc pondérés plus haut c'est-à-dire attribuer une importance. Les mots sont représentés par des vecteurs sémantique d'où la notion d'espace vectoriel VSM (Vector Space Model) [18].

### 2.3.2.2 Les mesures basées sur des connaissances :

Elle situe sur l'identification du degré de similitude entre les mots à l'aide d'informations dérivées de ressources externes tel que le réseau WordNet ou les dictionnaires de synonymes.[19] Dans la Figure 4 nous pouvons reconnaître quelques mesures les plus répondues.

### 2.3.3 Similarité hybride :

Les méthodes hybrides utilisent une combinaison de mesures de similarité, soit des méthodes sémantiques et syntaxiques à la fois où bien plusieurs méthodes de même type. Elle donne de supérieures valeurs du facteur de corrélation et donc de meilleurs résultats par rapport à l'application d'une seule mesure de similarité.[20]

## 2.4 Le Word embedding :

Le Word Embedding est un ensemble de techniques, ayant pour but le mappage de mots en vecteurs de nombres réels, dans un espace dimensionnel réduit. Cette méthode permet de capturer la similarité sémantique, ainsi que le contexte d'un mot dans un corpus donné.

Il existe plusieurs techniques de Word Embedding. Nous nous intéressons au Word2Vec, plus précisément au Skip-gram [21].

Pour l'arabe, nous avons utilisé le Word Embedding de Zahrane [21], qui comporte le mapping d'environ 6,3 million de mots en vecteur d'entier de taille 300.

Concernant l'anglais, nous nous sommes basés sur le Word Embedding faisant parti du référentiel de vecteurs de mots TALL [22]. Il comporte 4027169 mots représenté par des vecteurs de 100.

### 2.4.1 Continuous Bag Of Word(CBOW) :

L'architecture du modèle CBOW tente de prédire le mot cible actuel (le mot central) en fonction des mots de contexte source (mots environnants).D'autre façon La façon dont CBOW fonctionne c'est qu'elle tend à prédire la probabilité d'un mot donné un contexte.

Un contexte peut être un mot unique ou un groupe de mots La figure 5 montre l'architecture de ce modèle. Le but du réseau de neurones CBOW est de maximiser l'équation suivante : [23]

$$\frac{1}{V} \sum_{\tau=1}^V \log p(m_{\tau} | m_{\tau - \frac{c}{2}} \dots m_{\tau + \frac{c}{2}})$$

Où V (et C) est la taille du vocabulaire (contexte).

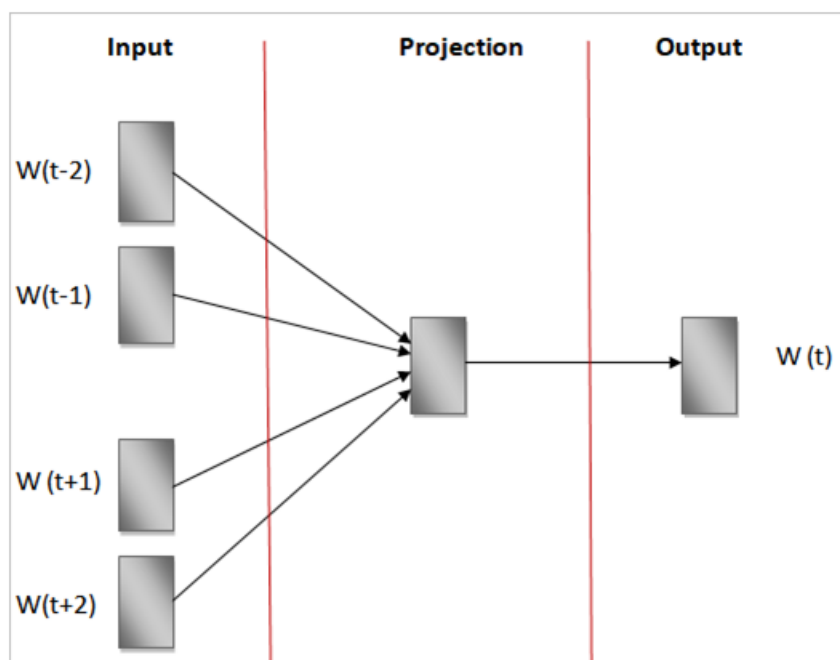


FIGURE 2.5 – L'architecture de modèle CBOW.

### 2.4.2 Skip-Gram(SG) :

L'objectif de formation du modèle Skip-Gram est de trouver des représentations de mots qui sont utiles pour prédire les mots environnants dans une phrase ou un document plus formellement. La figure 6 montre l'architecture de ce modèle. Le but du réseau de neurones SkipGram est de maximiser l'équation suivante : [23]

$$\frac{1}{V} \sum_{t=1}^V \sum_{j=t-c, j \neq t}^{t+c} \log p(m_j \setminus m_t)$$

Où  $V$  (et  $C$ ) est la taille du vocabulaire (contexte).

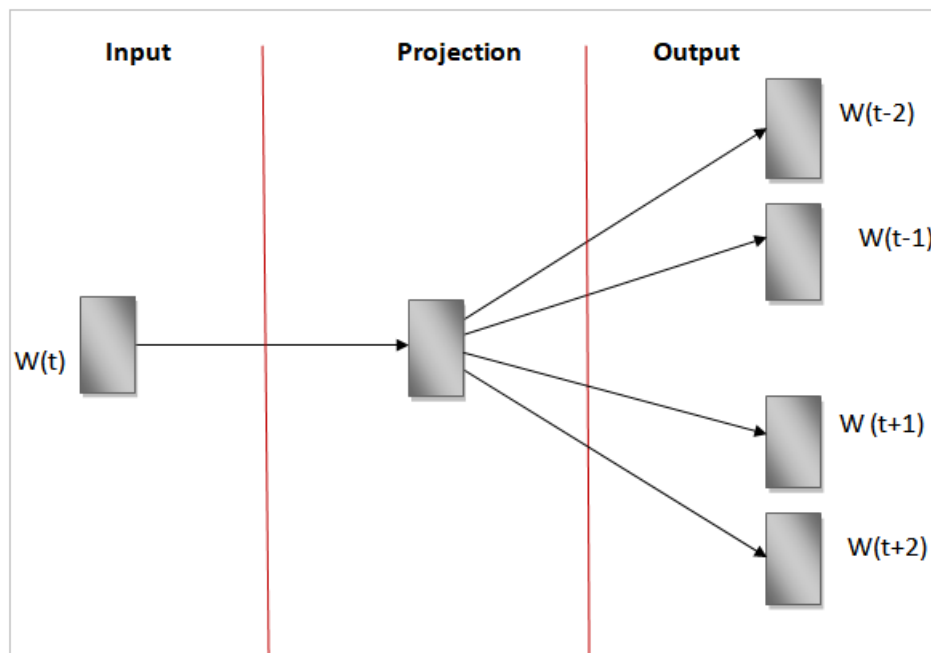


FIGURE 2.6 – L'architecture de modèle SkipGram.

## 2.5 Concepts fondamentaux liés à l'apprentissage automatique :

### 2.5.1 Définitions :

Le domaine de l'apprentissage automatique est parmi les domaines les plus connus dans le domaine de l'intelligence artificielle.

En 1959, Arthur Samuel a défini l'apprentissage automatique comme un « domaine d'étude qui donne aux ordinateurs la capacité d'apprendre sans être explicitement programmés. » (Swamynathan, 2017). Or que les chercheurs Borovcnik, Bentz, et Kapadia ont définies l'apprentissage automatique comme un ensemble de méthodes capables

de détecter automatiquement des modèles dans les données, puis d'utiliser les modèles découverts pour prédire les données futures ou pour effectuer d'autres types de prise de décision dans l'incertitude. (Borovcnik, al. 1991).

## 2.5.2 Types de l'apprentissage automatique :

Dans le domaine du Machine Learning (apprentissage automatique en français), il existe deux principaux types d'apprentissages : supervisés et non supervisés. La figure « Figure 7 » présente les deux différents types de l'apprentissage automatique les algorithmes liés à chaque type :

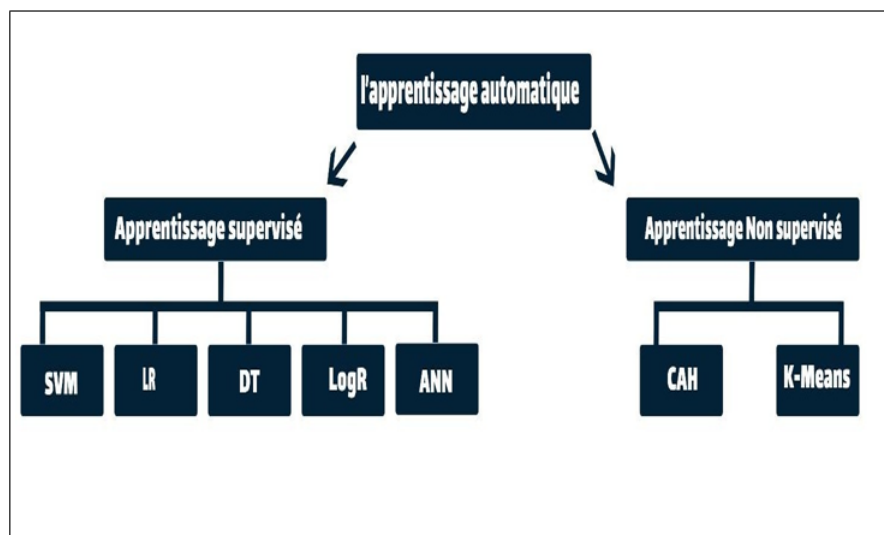


FIGURE 2.7 – les types d'apprentissage automatique avec leurs algorithmes.

### Apprentissage automatique non supervisé :

L'apprentissage non supervisé fait référence au processus de regroupement des données en grappes à l'aide de méthodes automatisées ou d'algorithmes sur des données qui n'ont pas été classées ou catégorisées. Dans cette situation, les algorithmes doivent « apprendre » les relations ou caractéristiques sous-jacentes à partir des données disponibles et regrouper les observations avec des caractéristiques similaires. (Berry et al., 2019)

Les problèmes d'apprentissage non supervisé peuvent être divisés en problèmes de regroupement (clustering) et d'association. Le clustering consiste à regrouper les points de

données par similarité ou distance. Connaître l'association nous permet de découvrir les règles qui décrivent la plupart des données.

Il existe deux principales méthodes d'apprentissage non supervisées :

- **K-moyen(k-means)** : L'algorithme K-Mean est l'un des algorithmes d'apprentissage non supervisé les plus simples pour résoudre les problèmes de clustering. Il est utilisé lorsque les données étiquetées ne sont pas disponibles.
- **Clustering ascendant hiérarchique (CAH)** : Selon (Chaitanya,2018) l'algorithme non supervisé CAH est une technique du regroupement( clustering ). On peut le résumer dans cette expression "Chaque point ou cluster est progressivement "absorbé" par le cluster le plus proche ".(Atif, 2016)

### 2.5.2.2 Apprentissage automatique supervisé :

L'apprentissage supervisé se concentre sur les modèles d'apprentissage en reliant la relation entre les variables et les résultats connus et en travaillant avec des ensembles de données étiquetés. L'apprentissage supervisé fonctionne en alimentant les données d'échantillon de la machine avec diverses fonctionnalités (représentées par «X ») et la sortie de valeur correcte des données (représentée par « y ») (Theobald 2017) ,la figure 7 représente la phase d'apprentissage.

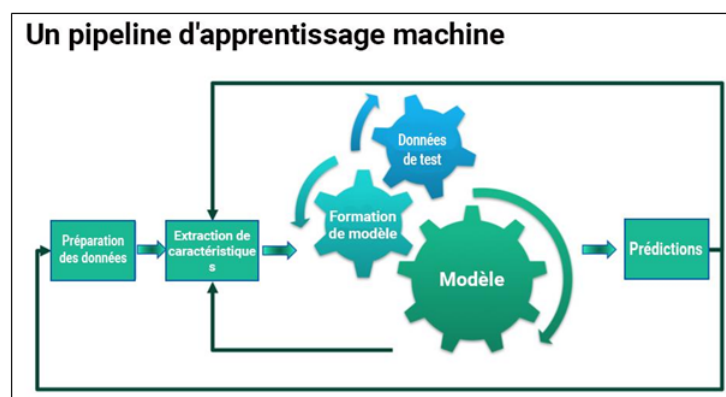


FIGURE 2.8 – Phase d'apprentissage du modèle d'apprentissage supervisé.

Il existe deux principales modèles d'apprentissage supervisées (Regression et Classification) , comme il est illustré dans la figure 9 :

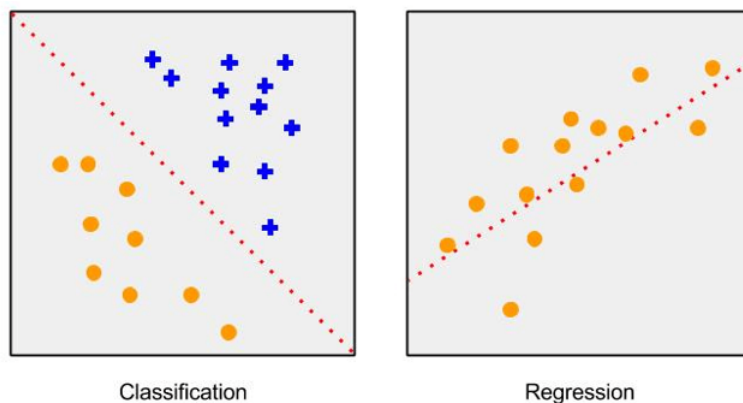


FIGURE 2.9 – Les différents problèmes de l'apprentissage automatique supervisé.

### a. Classification :

Lorsque les variables à prédire prennent des valeurs discrètes, on parle de problèmes de classification. L'algorithme de classification peut prédire une valeur continue, mais la forme de la valeur continue est la probabilité de l'étiquette de classe.

### b. Régression :

Lorsque la variable de sortie est une valeur réelle, un problème de régression se produit. Un algorithme de régression peut prédire une valeur discrète, mais la valeur discrète est un entier. Plusieurs algorithmes d'apprentissage automatique supervisé sont utilisés pour résoudre les problèmes de classification ou de régression dans ces algorithmes que nous avons :

- **Linéaire régression (Linear regression)** : La régression est principalement utilisée pour trouver la relation entre les variables et les prédictions, ou pour utiliser la valeur d'autres variables ( $x$ ) pour estimer la valeur attendue de l'équation.
- **Régression logistique (Logistic Regression)** : Selon (Ayush Pant, 2019), la régression logistique est un algorithme d'apprentissage automatique supervisé pour les problèmes de classification. C'est un algorithme d'analyse prédictive basé sur le concept de probabilité. Il est utilisé pour affecter des observations à un ensemble de classes discrètes.
- **Forêt aléatoire (Random forest)** : (Tony Yiu, 2019), La forêt aléatoire est un algorithme de classification composé de nombreux arbres de décision fonctionnant comme un ensemble. Il a utilisé l'ensachage et le hasard lors de la construction de chaque arbre in-

dividuel pour essayer de créer une forêt d'arbres indépendants. Les prédictions du comité étaient plus précises que celles de n'importe quel arbre individuel.

- **Machine à vecteurs de support (Support Vector Machine) :** C'est un classificateur discriminant défini formellement par un hyperplan de séparation. Autrement dit c'est un algorithme d'apprentissage automatique supervisé qui peut être utilisé pour des défis de classification. Dans cet algorithme, nous traçons chaque élément de données comme un point dans un espace à  $n$  dimensions (où  $n$  est le nombre d'entités qu'on a), la valeur de chaque entité étant la valeur d'une coordonnée particulière. Ensuite, nous effectuons la classification en trouvant l'hyperplan qui différencie très bien les 2 classes.

- **Arbre de décision (Decision tree) :** (Avinash Navlani, 2018), L'arbre de décision est un algorithme de traitement automatique supervisé. Il s'agit d'une structure arborescente similaire à un organigramme, dans laquelle les nœuds internes représentent des caractéristiques (ou attributs), les branches représentent des règles de décision et chaque nœud feuille représente un résultat.

- **Réseaux de neurones artificiels (Artificial Neural Networks-ANN) :** Un réseau de neurones artificiels ou ANN est un paradigme de traitement de l'information inspiré de la façon dont les systèmes nerveux biologiques (tels que le cerveau) traitent l'information. Il se compose d'un grand nombre d'éléments de traitement hautement interconnectés (neurones) qui travaillent ensemble pour résoudre des problèmes spécifiques. (Naqsh Singh Johan, 2019).

La figure suivante « Figure 10 » représente un modèle général d'un réseau de neurone artificiel qui est inspiré d'un neurone biologique.

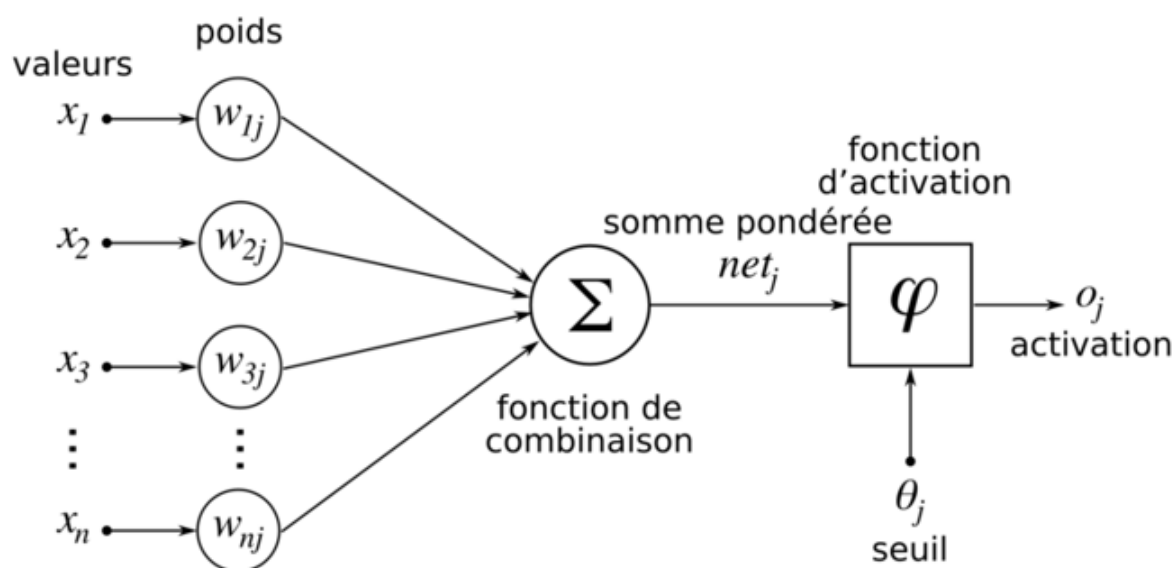


FIGURE 2.10 – Modèle général d'un réseau de neurone artificiel

## 2.6 Métriques d'évaluation des modèles :

La sélection et la construction d'un modèle de machine learning doivent passer par une étape d'évaluation pour mesurer ses performances et l'efficacité des résultats. Nous pouvons distinguer les problèmes de régression qui ont des métriques qui peuvent être utilisées pour évaluer les modèles d'apprentissage automatique et d'autres métriques qui sont utilisées pour les problèmes de classification :

**Précision de classification (Classification Accuracy) :** Il s'agit du rapport entre le nombre de prédictions valides et le nombre total d'échantillons d'entrée. Cela ne fonctionne bien que s'il y a un nombre égal d'échantillons appartenant à chaque classe. (Mishra, 2018)

$$Accuracy = \frac{\text{Number of Correct predictions}}{\text{Total number of predictions made}}$$

**Matrice de confusion (Confusion Matrix) :** (Kohavi et al., 1998) Comme son nom l'indique, la matrice de confusion nous fournit une matrice de sortie et décrit la performance complète du modèle (Mishra, 2018). Supposons que nous ayons un problème de

classification binaire. Nous avons des échantillons qui se répartissent en deux catégories : OUI ou NON. De plus, nous avons notre propre classificateur qui peut prédire la classe d'un échantillon d'entrée donné. En testant notre modèle sur 165 échantillons, nous obtenons les résultats suivants :

n=165	Predicted: NO	Predicted: YES
Actual: NO	50	10
Actual: YES	5	100

Il y a 4 termes importants :

**Vrais positifs** : Les cas dans lesquels nous avons prédit OUI et la sortie réelle étaient également OUI.

**Vrais négatifs** : Les cas dans lesquels nous avons prédit NON et la sortie réelle étaient NON.

**Faux positifs** : Les cas dans lesquels nous avons prédit OUI et la sortie réelle étaient NON.

**Faux négatifs** : Les cas dans lesquels nous avons prédit NON et la sortie réelle étaient OUI. La précision de la matrice peut être calculée en prenant la moyenne des valeurs situées sur la « diagonale principale », c'est-à-dire :

$$Accuracy = \frac{TruePositives + FalseNegatives}{TotalNumberofSamples}$$

Confusion Matrix constitue la base des autres types de mesures.

- **Rappel (Recall)** : (Kulhare, 2017), TP désigne la taille de l'ensemble positif vrai, FP désigne la taille de l'ensemble faux positif et FN désigne la taille de l'ensemble faux négatif, la précision et le rappel sont définis par :

$$\text{recall} = \frac{TP}{TP + FN}$$

**Précision :**

$$\text{precision} = \frac{TP}{TP + FP}$$

**F1 Score :** Le score F1 peut être interprété comme une moyenne pondérée de la précision et du rappel, où un score F1 atteint sa meilleure valeur à 1 et son pire score à 0. La contribution relative de la précision et du rappel au score F1 est égale :

$$F1 = 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

a) Métriques d'évaluation de modèles pour les problèmes de régression :

• **Erreur quadratique moyenne (Mean Squared Error) :**

L'erreur quadratique moyenne (MSE) est exactement la même que l'erreur absolue moyenne, la seule différence est que la MSE prend la différence quadratique moyenne entre les valeurs d'origine et les valeurs attendues. L'avantage de MSE est qu'il est plus facile de calculer le gradient, tandis que l'erreur absolue moyenne nécessite des outils de programmation linéaire préparés pour calculer le gradient. Lorsque nous prenons le carré de l'erreur, l'effet des erreurs plus importantes devient plus prononcé que l'erreur plus petite, donc le modèle peut désormais se concentrer davantage sur les erreurs plus importantes. (Mishra, 2018)

$$\text{MeanSquaredError} = \frac{1}{N} \sum_{j=1}^N (y_j - \hat{y}_j)^2$$

- **Erreur absolue moyenne (Mean Absolute Error) :**

(Willmott, et al . 2005) : L'erreur absolue moyenne est la différence moyenne entre les valeurs originales et attendues. Il nous donne une mesure de la distance entre les prévisions et la sortie réelle. Cependant, ils ne nous donnent aucune idée de la direction de l'erreur, c'est-à-dire si nous sommes sous ou surestimés. Mathématiquement, il est représenté par :

$$\text{MeanAbsoluteError} = \frac{1}{N} \sum_{j=1}^N |y_j - \hat{y}_j|$$

- **Le coefficient de corrélation de Pearson :**

(karl Pearson, 1895)(Wu et al.,2019) a défini le coefficient de corrélation comme suit : Le coefficient de corrélation de Pearson est généralement représenté par la lettre r, si nous avons un ensemble de données  $x_1, \dots, x_n$  contenant n valeurs et la prédiction de l'ensemble de données  $y_1, \dots, y_n$  contenant n valeurs , alors cette formule pour r est :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

Où n est la taille de l'échantillon,  $x_i$  est l'échantillon indexé avec i,  $y_i$  est la prédiction du système correspondant, et  $\bar{x}$ ,  $\bar{y}$  sont les moyennes de  $x_i$  et  $y_i$ , respectivement.

- **Erreur quadratique moyenne racine (Root Mean Squared Error) :**

Selon (Neill et al., 2018) L'erreur moyenne de racine carrée est la racine carrée de l'erreur quadratique moyenne de toutes les erreurs. L'utilisation du RMSE est très courante et est considérée comme une excellente mesure d'erreur à usage général pour les prédictions numériques. C'est une bonne mesure de précision pour comparer les erreurs de prédiction de différents modèles.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - O_i)^2}$$

Où  $O_i$  sont les observations,  $S_i$  les valeurs prédites d'une variable et  $n$  le nombre d'observations disponibles pour l'analyse.

- **Le kappa pondéré quadratique :**

Le Kappa pondéré mesure la cohérence entre deux factures, et cet indicateur varie généralement de 0 à 1. Si l'accord entre les évaluateurs est plus faible que prévu par hasard, l'indicateur peut tomber en dessous de zéro. (Arora aman, 2019) a présenté la méthode de calcul du kappa quadratique pondéré comme suit : Tout d'abord, une matrice d'histogramme  $N \times N$   $O$  est construite, de sorte que  $O_{i,j}$  correspond au nombre d'enregistrements d'adoption qui ont une note  $i$  (réelle) et ont reçu une note prédite  $j$ . Une matrice de pondérations  $N \times N$ ,  $w$ , est calculée en fonction de la différence entre les scores de notation réels et prévus.

$$\kappa = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}}$$

## 2.7 Les Revue de la littérature

### 2.7.1 Systèmes qui utilisent le Machine Learning

Plusieurs travaux ont adopté les techniques d'apprentissage automatique (ML) pour la résolution des problèmes de correction automatique des réponses courtes (ASAG), on présent quelques travaux :

Numéro d'article	Titre de l'article	Citation
1	Automated Short-Answer Grading Using Deep Neural Networks and Item Response Theory	Masaki Uto et al(2020)
2	Stacking Neural Network Models for Automatic Short Answer Scoring	Rian Adam Rajagede et al(2020)
3	A Short Answer Grading System in Chinese by Support Vector Approach	Wu et al (2019)
4	Auto Grader for Short Answer Questions	Patil et al(2018)
5	An automatic short-answer grading model for semi-open- ended questions	Zhang et al (2019)
6	Deep learning for Short Answer Scoring	Surya K et al (2019)
7	Automated Essay Grading using Machine Learning Algorithm	V. V.Ramalingam et al(2018)
8	Automatic short answer grading feedback using text mining methods	SUZEN, et al (2018)

FIGURE 2.11 – Tableau 1 :ASAG Systèmes qui utilisent le Machine Learning

### 2.7.2 Systèmes pour la langue Arabe

Dans le domaine éducatif en langue arabe nous avons constaté qu'il y a très peu de projets de recherche sur le classement des réponses courtes. La langue arabe est une langue connue par la richesse de son vocabulaire qui nécessite un travail approfondi pour l'étudier.

Pour cela la tâche de notation automatique en langue arabe est devenue difficile à traiter en raison du manque de ressources de base, il n'y a pas assez de corpus accessible au public pour le classement des réponses courtes en Arabe, manque d'outils de traitement

automatique de la langue. Tous ces difficultés seront une sorte de motivation pour investir dans le domaine de la notation automatique des réponses courtes, enrichir les outils de manipulation de cette langue et aussi la recherche dans ce domaine.

Parmi les investissements de notation automatique pour la langue Arabe il y a :

Numéro de l'article	Article	Citation
1	Deep neural network approach for Arabic community question answering	Ali Almiman et al(2020)
2	Deep Contextualized pairwise semantic similarity for Arabic language questions	Hesham Al-Bataineh et al(2019)
3	AEGD: Arabic essay grading dataset for machine learning	Bassam al-shargabi et al(2021)

FIGURE 2.12 – Tableau 2 :ASAG Systèmes qui utilisent le Machine Learning la langue arabe

## 2.8 Conclusion

La correction automatique pour les réponses courtes, comme nous venons de le voir, est un domaine enraciné dans de nombreux autres domaines tels que le traitement automatique du langage, l'arithmétique de similarité et l'évaluation automatique.

Toutes ces ressources seront mobilisées pour développer la prochaine étape qui dépendra fortement de ce qui a été traité afin de concevoir notre système et construire une méthodologie basée sur les forces dérivées de cette vision et travaillait sur les faiblesses à traiter est capable de corriger-les plus tard. Dans l'étape suivante, nous présentons la méthodologie proposée pour construire notre système avant de commencer la phase de mise en œuvre et de discuter des résultats.

# Systeme d'évaluation automatique des réponses courtes

## **3.1 Introduction**

Cette partie est dédiée au système que nous voulons développer et utilise une méthode supervisée d'apprentissage automatique pour évaluer automatiquement les réponses courtes. Nous présenterons notre conception détaillée et les mécanismes qui y sont utilisés, tout en étudiant l'impact de l'apprentissage automatique sur l'amélioration des résultats précédemment obtenus pour l'approche ASAG. Nous avons présenté la méthodologie avec laquelle nous allons construire le système, puis, dans la section 4, nous avons discuté du plan suivi de la mise en œuvre du système.

## **3.2 Méthodologie**

Le système que nous mettons en œuvre essaie de bénéficier de la puissance et de l'efficacité de la machine learning, par exemple la qualité de ce dernier et l'obtention de résultats compétitifs dépendent des caractéristiques et des types de modèles d'ingénierie.

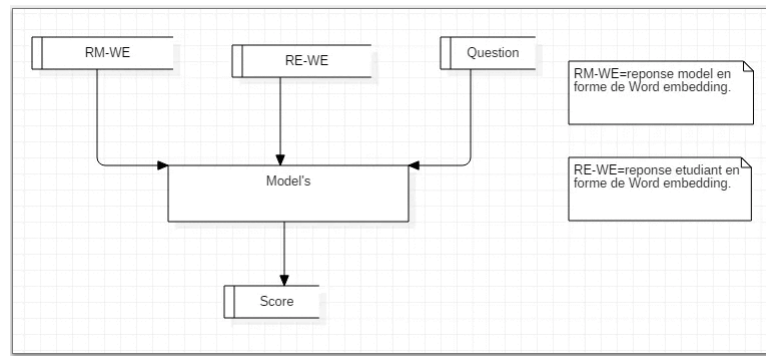


FIGURE 3.1 – Plan suivi dans la conception

### 3.3 Les modèles proposés

Le problème de notation automatique peut être considéré comme un problème de régression. Nous prédisons le score en fonction d'un ensemble de caractéristiques calculées ou de problèmes de classification. Nous assignons les étudiants aux classes appropriées en fonction de leurs notes. , Parce que le machine learning supervisé résout ces deux types de problèmes. Nous choisissons les modèles suivants afin de pouvoir choisir les modèles qui nous apportent des résultats significatifs :

La Régression : **LR**, **SVR**, **ANN**, **DTR**

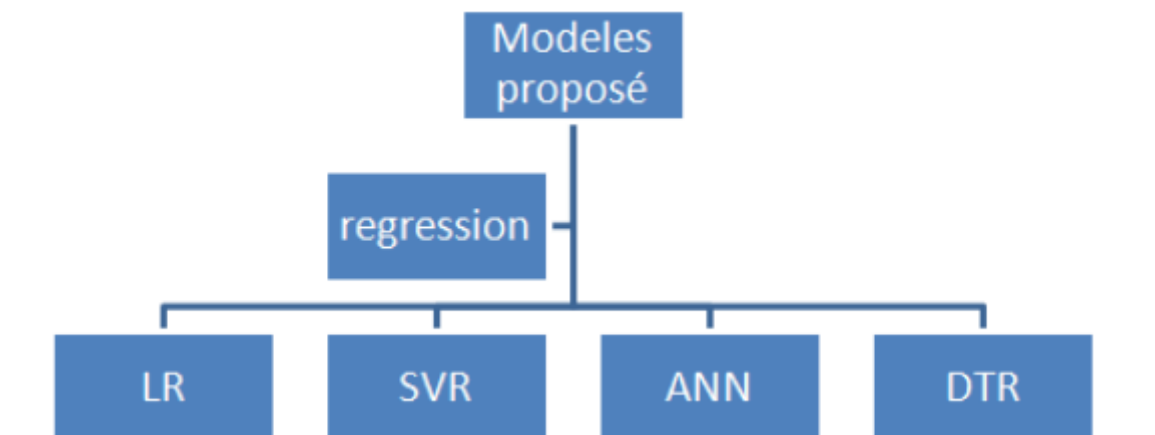


FIGURE 3.2 – Représentation des Modèles proposés

## 3.4 Mise en œuvre des modèles

Sur la base de la méthode ci-dessus, la mise en œuvre de notre système prendra plusieurs étapes pour atteindre le modèle final, ce qui nous permet de prédire automatiquement le score. Ces étapes sont expliquées en détail ci-dessous :

### 3.4.1 Prétraitement de Données :

La partie préparation des données est la partie la plus importante du processus de traitement du langage naturel. Il est nécessaire de nettoyer les données textuelles pour mettre en évidence les attributs que nous souhaitons que le système d'apprentissage automatique récupère. Le nettoyage (ou prétraitement) des données comprend généralement plusieurs étapes, telles que la suppression des mots vides, la ponctuation, la tokenisation et la normalisation. Ces étapes sont illustrées à la figure "Figure 13"

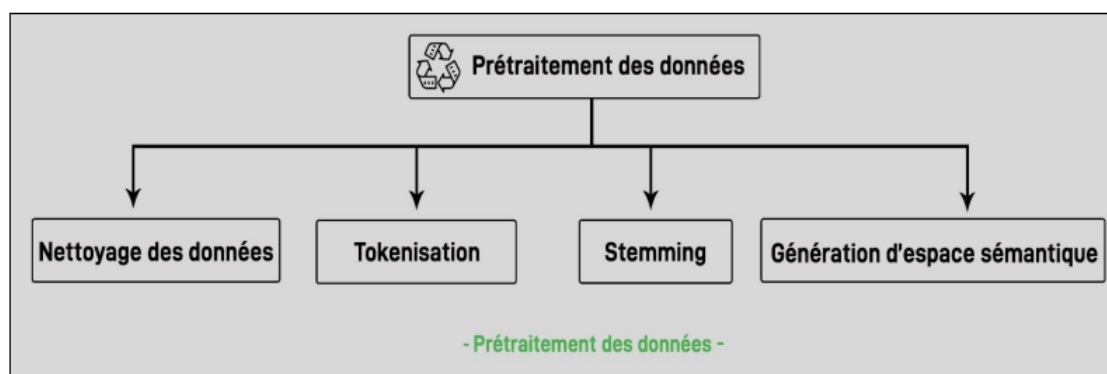


FIGURE 3.3 – Représentation des différentes étapes de prétraitement des données

#### Prétraitement de Données :

L'étape de nettoyage des données est la première étape de base de notre système, qui se compose de deux parties principales : « mots vides » et « supprimer la ponctuation ». Après l'étape de nettoyage des données, nous procéderons à la même étape importante que la première étape, à savoir la tokenisation.

**Tokenisation :**

La Tokenisation consiste à découper une séquence de caractères phrase, paragraphe ou un document texte entier en morceaux, appelés jetons.

**Exemple :** la phrase avant et après la tokenisation :

نشر الديمقراطية سيؤدي الى السلام  
نشر, الديمقراطية, سيؤدي, الى, السلام

Nous utilisons la tokenisation pour plusieurs raisons : il est facile de calculer la similarité, car nous utiliserons le vecteur de contexte de chaque terme, utiliserons différentes méthodes de similarité syntaxique pour les termes, et enfin l'utilisation de Stemming, dont nous parlerons dans la section suivante.

**Stemming :**

Stemming fait partie des études linguistiques sur la recherche et l'extraction d'informations sur la morphologie et l'intelligence artificielle (IA). Les connaissances sur les racines et l'IA extraient des informations significatives à partir de vastes sources telles que les méga données ou Internet, car des formes supplémentaires d'un mot liées à un sujet peuvent nécessiter une recherche pour obtenir les meilleurs résultats. Stemming fait également partie des requêtes et des moteurs de recherche Internet.

.La manipulation des textes est une approche très sensible surtout pour les mesures de similarités, un caractère de plus peut faire la différence, c'est pour cela nous pensons à utiliser le Stemming, et il a prouvé sa performance.

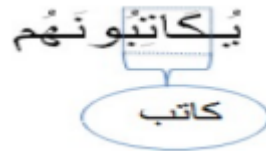
**Types de stemming :** il existe deux types de stemming le« **stemming léger-light stemming** » , et le« **stemming lourd- root stemming** » pour extraire la tige des mots.

les deux types a but du stemming est de supprimer l'affixe (préfixes, infixes et suffixes)

d'une chaîne, en renvoyant la tige du mot en sortie.

**Stemming lourd** : Le stemming lourd, également appelé « Root-Stemming » (Stemming à la racine), consiste à supprimer les préfixes et les suffixes bien connus pour extraire la racine réelle d'un mot et à identifier le motif en correspondance avec le mot restant.

**Le stem lourd (Heavy Stemming)** : son objectif est d'extraire le Stem et la racine.



**Stemming léger** : Le stemming léger est un processus moins complexe, où le stemming est arrêté sur la suppression des préfixes et des suffixes, sans tenter d'identifier la racine réelle du mot.

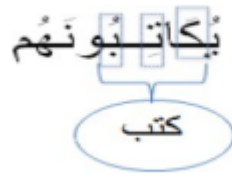
De nombreux stemmers légers ont été recommandés pour la Langue arabe [25].

Pour atténuer l'impact de l'inconvénient majeur des algorithmes basés sur la racine, qui fait perdre la notion de sémantique, une lemmatisation légère provenant de l'arabe a également été proposée. Stemmers légers coupent quelques affixes tels que les terminaisons plurielles en anglais à partir des mots et sans effectuer d'analyse linguistique profonde.

De ce point de vue, la majorité des approches tentent de supprimer les préfixes les plus fréquents (c'est-à-dire articles), les suffixes (c'est-à-dire les pronoms possessifs) et tous les préfixes ou suffixes cela peut être attaché au début ou à la fin des mots.

Par exemple, les stemmers légers génèrent (ce qui signifie : œuvres) de parce que seuls les préfixes (y compris les antéfixes) et les suffixes (y compris les post fixes) sont supprimés. La décision de retirer cependant, les affixes sont généralement contrôlés par des règles heuristiques dérivées de l'utilisation commune de ces antéfixes [26].

**Le stem léger (Light Stemming)** : s'intéresse seulement par le Stem.



Parmi les stemmers léger, nous mentionnons :

**Tashaphyne stemmer** Tashaphyne est un stemmer arabe de poids léger (Zerrouki, 2010) [27]. Il commence par supprimer les lettres non liées dans le processus d'extraction de racine, tels que les signes diacritiques. Il utilise deux listes de préfixes et suffixes pour segmenter un mot donné. Tashaphyne fournit à la fois une tige légère ou une racine au mot d'entrée.

**Les stemmers utilisés :** Dans notre travail nous utilisons différent stemmer connus en langue qui sont :

- Isri Stemmer (pour un stemminglourd)
- Tashapyne (pour un stemming lourd etléger)
- PorterStemmer (Pourl'anglais)

**Génération d'espace sémantique :** Dans cette étude nous essayons de bénéficier des outils déjà construits comme les générations d'espaces sémantiques développés par les étudiants d'université de Blida en utilisant la méthode **COALS**.

## 3.5 Les Etapes d'élaboration d'espace sémantique :

### 3.5.1 Collection des informations spécifiques de domaine :

L'information de domaine que nous voulons étudier est considérée comme le matériau principal pour l'établissement de l'espace sémantique, nous l'extrayons donc d'Internet, de livres ou d'autres ressources qui se concentrent sur le même domaine afin de les convertir en texte et de les structurer en fichiers texte pour un traitement facile.

### 3.5.2 Prétraitement de corpus :

Une fois les documents textuels du corpus de texte collectés, des préparatifs doivent être faits pour que ces derniers soient traités plus avant. Cette étape est basée sur un ensemble d'étapes de prétraitement pour éliminer certaines ambiguïtés, réduire la quantité de traitement ultérieur et adapter le corps du texte à l'objectif final de « l'extraction du mot candidat ». (Il s'agit d'utiliser des outils de traitement automatique en arabe pour nettoyer le document afin de sélectionner le meilleur stemmer).

### 3.5.3 Génération des termes uniques :

Permet de représenter des textes sans aucune distorsion au niveau des caractères lors de la lecture, nous allons trier de manière unique chaque mot et le placer dans un fichier de sorte que le numéro de ligne de chaque mot de ce fichier soit un index de chaque ligne de la matrice de vide sémantique.

### 3.5.4 Elaboration de Matrice d'espace sémantique :

Dans la dernière étape, nous calculons les fréquences pour chaque mot dans le fichier des mots uniques en utilisant la méthode **COALS** à partir de leur apparition dans le fichier du domaine étudié, que nous avons créé dans la deuxième étape.

	البرامج	السرقه	أمن	بيانات	الهاكر	الهندسة	جرائم	المستخدم	الانترنت
البرامج	0	0.23	0.06	0.1	0.62	0.01	0.33	0.2	0.9
السرقه	0.23	0	0.51	0.66	0.15	0.8	0.9	0.4	0.92
أمن	0.06	0.51	0	0.8	0.07	0.002	0.9	0.001	0.3
بيانات	0.1	0.66	0.8	0	0.9	0.51	0.01	0.66	0.23
الهاكر	0.62	0.15	0.07	0.9	0	0.06	0.62	0.4	0.15
الهندسة	0.01	0.8	0.002	0.51	0.06	0	0.15	0	0.51
جرائم	0.33	0.9	0.9	0.01	0.62	0.15	0	0.07	0.4
المستخدم	0.2	0.4	0.001	0.66	0.4	0	0.07	0	0.15
الانترنت	0.9	0.92	0.3	0.23	0.15	0.51	0.4	0.15	0

FIGURE 3.4 – Matrice d'espace sémantique

### 3.6 L'approche Machine Learning :

L'Apprentissage automatique est l'aspect clé de tous les caractéristiques présentées précédemment comme caractéristiques des textes, caractéristiques déviation-question et les caractéristiques de domaine général et spécifiques pour la résolution du problème de la notation automatique dont le but est de prédire un score ou bien faire une classification sur une échelle de notes adéquate.

Après l'étape de prétraitement et calcul des similarités en créant l'ensemble de données qui représente les caractéristiques sur lesquelles le modèle final est construit, nous enchaînons la dernière partie qui est la partie d'apprentissage automatique d'où nous parlons sur le partitionnement des données en données d'entraînement et données de teste (**Data Split**) et la combinaison des fonctionnalités pour l'obtention de meilleur modèle.

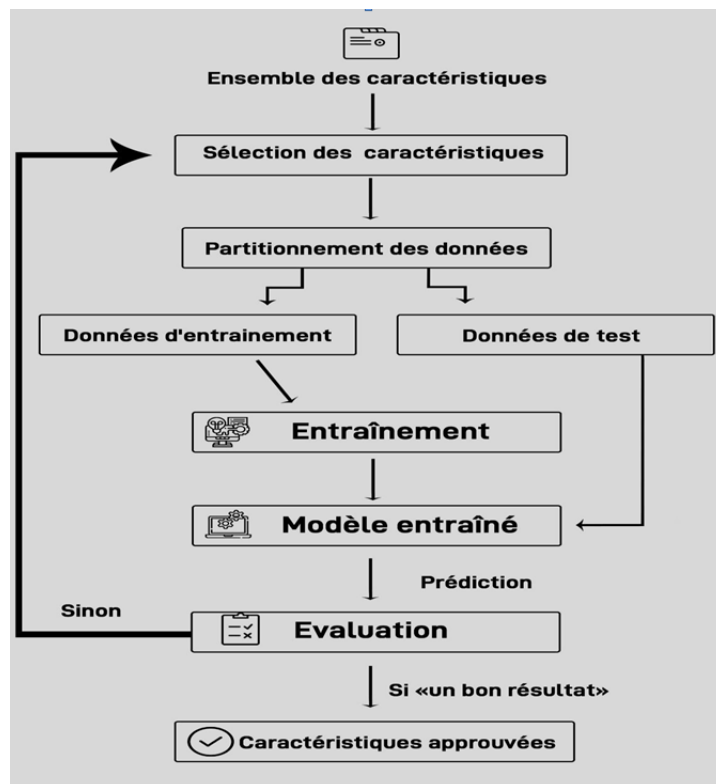


FIGURE 3.5 – Mécanisme suivi dans l'apprentissage automatique de notre système

Nous expliquons le mécanisme de ce processus dans la partie suivante.

### 3.6.1 Sélection des caractéristiques :

Pour étudier l'effet des caractéristiques sélectionnées sur le score de l'étudiant qui donne des résultats plus proches des résultats manuels, nous nous appuyons sur la création d'une harmonie entre les caractéristiques afin d'obtenir le meilleur résultat en termes de précision. La moindre différence entre la note manuelle et la note automatique.

Cette procédure est effectuée afin de préserver les ensembles de caractéristiques qui conduisent notre modèle à obtenir la plus grande précision.

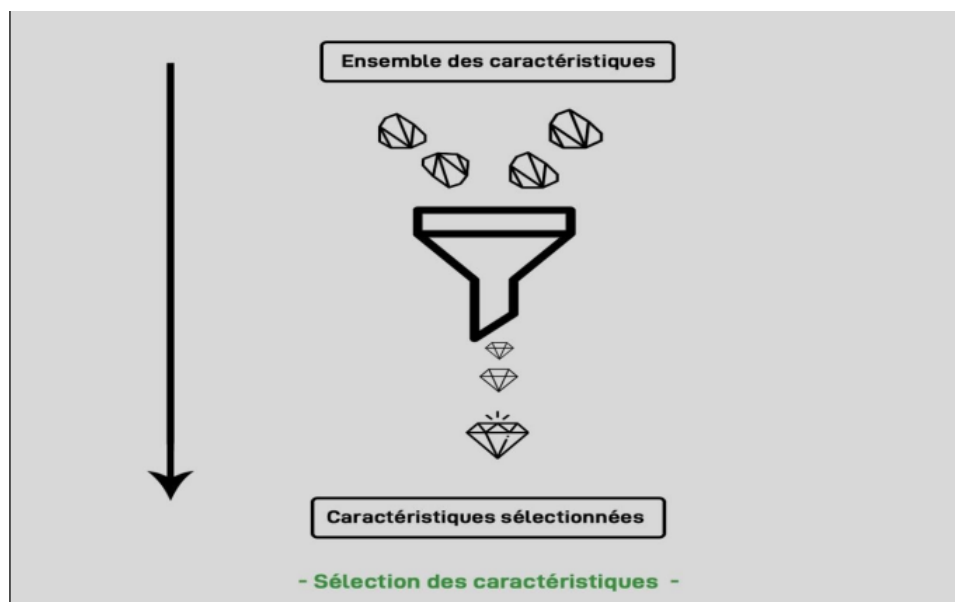


FIGURE 3.6 – Processus de l'extraction des caractéristiques

### 3.6.2 Partitionnement des données :

La segmentation des données a un impact très important dans les modèles d'apprentissage automatique, c'est donc une étape essentielle. Le fractionnement des données consiste à diviser l'ensemble de données en deux parties « données d'entraînement » et « données de test » ou en trois parties « données d'entraînement », « données de vérification » et

« données de test ». Test.” La partie entraînement a autant de couple que la partie test pour un bon entraînement pour un meilleur résultat.

### **3.6.3 Entraînement du modèle :**

L'étape d'apprentissage consiste à ajuster les paramètres de l'algorithme choisi en apprenant à partir de l'ensemble de données et en essayant d'atteindre les facteurs de propriétés qui conduisent à des résultats proches des résultats réels.

### **3.6.4 Test du modèle :**

Dans la partie test du modèle, nous prenons les données de test comme entrée dans l'algorithme afin de prédire les scores et de vérifier les performances et l'efficacité de l'algorithme.

### **3.6.5 Evaluation :**

Une fois que le modèle a été formé et testé, il passe par une étape d'évaluation qui est une étape essentielle dans la validation des propriétés qui conduit à un meilleur algorithme. L'évaluation est effectuée à l'aide de diverses mesures informatiques de la performance du modèle.

## **3.7 Conclusion**

Ce chapitre traite des méthodes utilisées dans notre ingénierie des systèmes et de notre conception finale utilisant cette idée.

Ci-dessous, nous présenterons la technologie utilisée dans l'application et présenterons la méthode de mise en œuvre avec les résultats.

# Résultats expérimentaux et évaluation

## 4.1 Introduction

Dans ce chapitre, nous présentons les expériences et les résultats des tests réalisés afin de voir l'amélioration des résultats par rapport aux travaux déjà réalisés, et discutons ainsi des résultats obtenus pour montrer l'effet de la construction de l'ensemble de réponses du modèle.

## 4.2 Implémentation :

La partie implémentation comprend tous les processus impliqués lors de la réalisation de notre système final tels que l'ensemble de données utilisé ainsi que sa structure, les techniques de prétraitement, l'extraction de fonctionnalités, les processus d'apprentissage automatique et nous parlons également de l'environnement de développement.

### 4.2.1 DataSet :

Dans notre système, nous utilisons le « ensemble de données AR-ASAG » fourni par (F.R. OUKINA et al, 2019) comme ensemble de données principal pour l'entraînement et le test des modèles proposés. Cet ensemble de données couvre les cours du module Cybercriminalité. Notre ensemble de données contient 48 questions, dont 48 réponses typiques et un ensemble de réponses d'élèves avec un total de 2 133 réponses avec des notes manuelles comprises entre 0 et 5 évaluées par deux enseignants.

Pour examiner les performances et la fiabilité de notre modèle, nous l'avons également testé sur un autre ensemble de données « Mohler-Dataset » (Mohler et al, 2011) qui contient 87 questions et un total de 2442 réponses d'étudiants.

Le tableau « Tableau 4.1 » est un récapitulatif de ce que nous avons présenté ci-dessus :

DATASET	NOMBRE DE QUESTIONS	NOMBRE DE REPONSES MODELE	NOMBRE DE REPONSES ETUDIANTS
LYCEE_ELASNAM	10	10	360
AR-ASAG	48	48	2133
MOHLER	87	87	2442

4.1.1

4.1.2 Tableau : Informations sur les Dataset utilisées

FIGURE 4.1 – Tableau 4.1 dataset

## 4.2.2 Structuration de DataSet :

Pour standardiser notre travail, le rendre global et facile pour l'exploitation, nous structurons chaque dataset en deux fichiers **XML**, de sorte que les réponses des étudiants sont dans un seul fichier tel que chaque réponse à un « **Question\_ID** » qui indique l'identifiant de la question, « **Answer\_ID** » qui indique l'identifiant de la réponse d'une question et « **Average\_Mark** » qui indique la note manuelle de l'étudiant.

La figure 4.1 montre la structure du premier fichier XML de l'ensemble de données **AR-ASAG Dataset** :

```

    <Average_Mark>4.0</Average_Mark>
  </Answer>
  <Answer>
    <Answer_ID>5</Answer_ID>
    <Question_ID>1.1</Question_ID>
    <Answer_Arabic>الالكترونية (الهاتف، الكمبيوتر...) ، يتمثل في حصول مي كل سلوك غير أخلاقي يتم باستخدام الوسائل
    <Average_Mark>3.5</Average_Mark>
  </Answer>
  <Answer>
    <Answer_ID>6</Answer_ID>
    <Question_ID>1.1</Question_ID>
    <Answer_Arabic>غالبا إلتلاف أو سرقة المعلومات و هي متلا هي سلوك غير قانوني عبر أجهزة إلكترونية، لأهداف مادية
    <Average_Mark>3.75</Average_Mark>
  </Answer>
  <Answer>
    <Answer_ID>7</Answer_ID>
    <Question_ID>1.1</Question_ID>
    <Answer_Arabic>جرم منه على فوائد مادية و معنوية، يتحمل ،هي سلوك غير قانوني يتم باستخدام الأجهزة الإلكترونية
    <Average_Mark>4.5</Average_Mark>
  </Answer>
  <Answer>
    <Answer_ID>8</Answer_ID>
    <Question_ID>1.1</Question_ID>
    <Answer_Arabic>ة و معنوية مع تحميل الضحية خسارة مقابلة تتم بواسطة الأجهزة الإلكترونية ينتج عنها حصول المجرم
    <Average_Mark>5.0</Average_Mark>
  </Answer>

```

FIGURE 4.2 – fichier XML qui représente l’ensemble de données « AR-ASAG Dataset ».

Le deuxième fichier contient tous les questions avec leurs réponses modèles telles que chaque question a un « **Question\_ID** ».

La figure 4.2 contient la représentation des questions de l’ensemble de données « AR-ASAG »

```
<?xml version="1.0" encoding="UTF-8" standalone="no"?>
<dataroot>
  <Question>
    <Question_ID>1</Question_ID>
    <Question_Type>1</Question_Type>
    <Question_Arabic>عرف مصطلح الجريمة الإلكترونية</Question_Arabic>
    <Model_Arabic>الإلكترونية) الهاتف، الكمبيوتر، الانترنت (ينتج عنه أي كل سلوك غير قانوني يتم باستخدام الأجهزة</Model_Arabic>
  </Question>
  <Question>
    <Question_ID>2</Question_ID>
    <Question_Type>1</Question_Type>
    <Question_Arabic>عرف مصطلح أمن المعلومات</Question_Arabic>
    <Model_Arabic>لمعلومات من منشآت نفسها والأفراد العاملين فيها حماية وتأمين كافة الموارد المستخدمة في معالجة</Model_Arabic>
  </Question>
  <Question>
    <Question_ID>3</Question_ID>
    <Question_Type>1</Question_Type>
    <Question_Arabic>عرف مصطلح الهندسة الاجتماعية النفسية</Question_Arabic>
    <Model_Arabic>خداع الآخرين والحصول على معلومات سرية يتم من أي ممارسة تستخدم أساليب التلاعب النفسي (العقلي)</Model_Arabic>
  </Question>
  <Question>
    <Question_ID>4</Question_ID>
    <Question_Type>1</Question_Type>
    <Question_Arabic>عرف مصطلح تبييض أو غسل الأموال</Question_Arabic>
    <Model_Arabic>ة غالبا من التجارة غير المشروعة في المخدرات أو غيرها يتوطينها في نشاطات ومعية على الانترنت</Model_Arabic>
  </Question>
</Question>
```

FIGURE 4.3 – fichier XML qui représente « Questions-Réponses Modèle AR-ASAG Dataset ».

### 4.2.3 Les techniques utilisées dans le prétraitement :

#### Normalisation :

Plusieurs étapes de nettoyage et de normalisation du corpus combiné ont été appliquées pour le texte collecté nous notons :

- Nettoyage des caractères bruyants, des balises et suppression des signes diacritiques.
- Normalisation de tous les chiffres numériques.
- Normalisation des caractères arabes :

(آ, إ, أ) à (ا) et (ة) à (ه).

#### Exemple :

#### StopWords :

une base de données qui contient **841mots** qui sont considérés comme des mots d'arrêt pour la langue arabe comme :

تبييض الأموال، هو إضفاء صفة المشروعية على الأموال القذرة المتأتية غالبا من التجارة غير المشروعة مثل: (المخدرات \_ بيع الأشياء المسروقة \_ تزوير العملات ...)، بتوظيفها في نشاطات وهمية على الأنترنت.

la normalisation

تبييض الاموال هو اضعاء صفه المشروعيه على الاموال القذره المتاتيه غالبا من تجاره غير المشروعة مثل المخدرات بيع الاشياء المسروقه تزوير العملات بتوظيفها في نشاطات وهميه على الانترنت

FIGURE 4.4 – Exemple de normalisation

« ... على , هو , ي ه »

Pour l'anglais, nous utilisons l'ensemble des **stop Word** de la bibliothèque **NLTK** qui contient **179 mots** comme : « the, or, and, to ... »

Exemple :

تبييض الاموال هو اضعاء صفه المشروعيه على الاموال القذره المتاتيه غالبا من تجاره غير المشروعة مثل المخدرات بيع الاشياء المسروقه تزوير العملات بتوظيفها في نشاطات وهميه على الانترنت

Suppression StopWord

تبييض الاموال اضعاء صفه المشروعيه الاموال القذره المتاتيه غالبا تجاره المشروعة المخدرات بيع الاشياء المسروقه تزوير العملات بتوظيفها نشاطات وهميه الانترنت

FIGURE 4.5 – Exemple de StopWord

Stemming :

nous utilisons le stem léger et lourd (**Isri, Tashaphyne**(lourd – léger) pour la langue arabe et **Porter Stemming, SnowBall** pour l'anglais.

**Espace sémantique :**

Nous générons la matrice de l'espace sémantique sur la base d'un ensemble spécifique de cours sur la cybercriminalité. Nous faisons de nombreuses expériences en créant la matrice de mots de différentes tailles, 9000 mots, 20000 mots et 30000 mots pour l'arabe. Pour Mohler, un ensemble spécifique de 15 000 et 30 000 mots est obtenu sur le Web.

**4.2.4 Extraction des caractéristiques :**

Cette étape regroupe tous les caractéristiques essentielles de notre système, nous parlons sur les techniques et les algorithmes du calcul de similarités utilisées, aussi sur les différentes pondérations appliquées dans ces calculs, et l'extraction de la difficulté de question en fonction des notes des étudiants.

**a) Caractéristiques du domaine général et spécifique :**

Les caractéristiques extraites des domaines généraux et/ou spécifiques sont le résultat de calculs de similarité sémantique, dans lesquels nous utilisons deux aspects de base principaux, qui sont importants dans le calcul de la similarité entre les réponses des élèves et les réponses du modèle.

La première méthode consiste à utiliser **l'espace sémantique**, qui est une technique basée sur un corpus spécifique, c'est-à-dire l'«corpus réseau ».

La deuxième méthode est **Word embedding** (la méthode d'incorporation de mots) . Dans cette partie, nous fournissons une explication de l'incorporation de mots. Nous discutons des techniques qui peuvent être utilisées pour apprendre des mots incorporés, des techniques utilisées dans notre système et des incorporations de nos mots préformés.

L'incorporation d'un mot est une représentation apprise d'un texte où les mots qui ont la même signification ont une représentation similaire. Ces mots sont représentés comme des vecteurs à valeur réelle dans un espace vectoriel prédéfini. Tout au long de notre travail nous intéressons sur la technique **Word2Vec** que nous utilisons pour l'approche des incorporations de mots d'où **Word2Vec** est une méthode statistique pour apprendre efficacement un mot autonome incorporé à partir d'un corpus de texte.

Deux architectures ont été initialement proposées pour apprendre les **Word2vec**, le modèle de sacs de mots continus (**CBOW** : continuous bag of words) et le modèle **skip-gram**.(Mikolov, et al., 2013)

Le modèle **CBOW** apprend l'intégration en prédisant le mot actuel en fonction de son contexte. Le modèle **skip-gram** apprend en prédisant les mots environnants à partir d'un mot courant.

La figure « Figure 4.6 » nous illustre les deux architectures des incorporations des mots : En pratique, le modèle CBOW est plus rapide à apprendre, mais le modèle skip-gram donne généralement de meilleurs résultats. (Mikolov, et al., 2013)

L'incorporation de mots est une représentation apprise du texte, dans laquelle les mots ayant le même sens ont des représentations similaires. Ces mots sont représentés sous forme de vecteurs à valeur réelle dans un espace vectoriel prédéfini. Tout au long de notre travail, nous nous sommes intéressés à la technologie Word2Vec pour les méthodes d'intégration de mots, donc Word2Vec est une méthode statistique pour apprendre efficacement des mots autonomes fusionnés à partir du corps du texte.

Il a été initialement proposé pour l'apprentissage de Word2vec, à savoir le modèle de sacs de mots continu (CBOW : continuous bag of words)) et le modèle de skip-gram (Mikolov, et al., 2013).

Le modèle CBOW apprend à intégrer en prédisant le mot courant en fonction du contexte. Le modèle skip-gram apprend en prédisant les mots environnants à partir d'un mot commun.

Notre système utilise deux incorporations de mots préformés le premier est **Word Embedding de zahran** , Une grande quantité de textes arabes bruts provenant de«**Wikipedia arabe** », «**Corpus Gigaword arabe** » et d'autres sources est collectée dans le système de (Mohamed A. Zahran, et al., 2015) . Le nombre de mots dans «**Word Embedding de zahran** » est plus de **6 millions mots**.

En pratique, le modèle CBOW est plus rapide à apprendre, mais le modèle skip-gram donne généralement de meilleurs résultats. (Mikolov, et al., 2013)

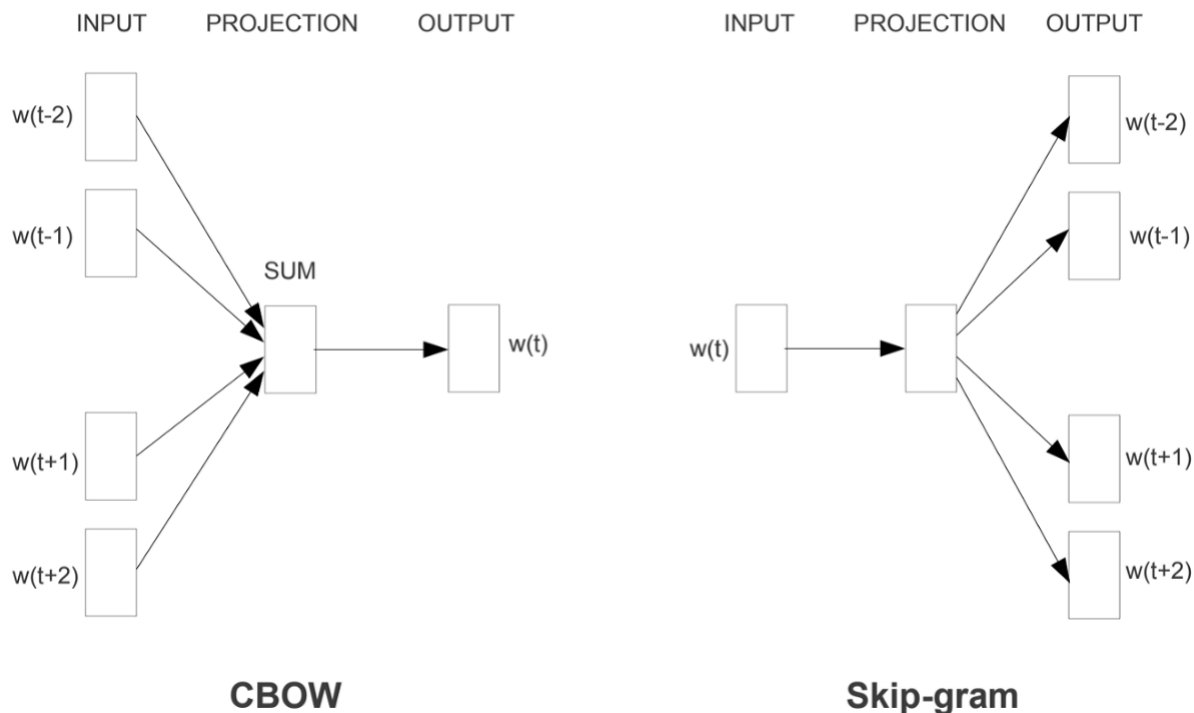


FIGURE 4.6 – Architectures de modèles CBOW et skip-gram

Dans notre travail nous exploitons des incorporations de mots préformés, afin que nous puissions les utiliser dans nos calculs.

Notre système utilise deux incorporations de mots préformés le premier est **Word Embedding de zahran**, Une grande quantité de textes arabes bruts provenant de «**Wikipedia arabe** », «**Corpus Gigaword arabe** » et d'autres sources est collectée dans le système de (Mohamed A. Zahran, et al., 2015) . Le nombre de mots dans «**Word Embedding de zahran** » est plus de **6 millions mots**.

Nous présentons un tableau « Tableau 4.2 » qui contient une représentation des vecteurs de contexte des mots

Ces vecteurs de contexte peuvent être pondérer en utilisant des aspects étudié ci-dessus comme suit :

Mots	Vecteurs
المعلومات	[0.582, 0.322, 0.423, 0.27, 0., 0., 0., 0., 0., 0.]
المخترق	[0.274, 0.423, 0., 0.33, 0.226, 0., 0., 0., 0., 0.]
المستخدم	[0., 0.27, 0.33, 0., 0.316, 0.21, 0., 0., 0., 0.]
البرامج	[0., 0., 0.226, 0.316, 0., 0.316, 0.226, 0.084, 0., 0.]

FIGURE 4.7 – Tableau 4.2 : Représentation vectorielle des mots

### La Pondération des termes :

Dans notre système, nous utilisons la méthode de pondération **TF-MinMax**, qui est générée à partir du corpus du domaine de recherche et d'une partie des balises vocales **PosTag**. Nous utilisons "**StanfordCoreNlp**" comme représentation de mot, puis nous pouvons attribuer des poids à chaque type d'étiquette afin que nous puissions les utiliser comme poids dans les calculs de similarité.

On trouve généralement dans la littérature que les verbes ont une grande influence sur le sens des phrases, c'est pourquoi nous attribuons aux verbes le poids le plus élevé, à savoir 0,5, les noms 0,3 et les autres 0,2. La somme de tous ces poids est égale à 1. Verbe + nom + autre = 1.

Nous utilisons la pondération TF-MinMax ou PosTag pour pondérer chaque vecteur de contexte (un vecteur obtenu par ES ou WE) de mots présents dans la réponse de l'élève avec un poids approprié.

### b) Caractéristiques des textes :

Parmi les fonctionnalités de texte, il existe de nombreuses technologies syntaxiquement similaires, qui sont décrites en détail dans la section "Concepts de base liés à la similarité de texte court". Nous avons utilisé les techniques suivantes :

- Basé sur des termes : similitude d'ordre, dé, calculs matriciels.
- Basé sur les caractères : similarité Jaccard, distance de **Levenshtein**, distance

de Jaro, chaîne universelle longue pour enfants (LCS), similarité STS.

### c) Caractéristiques d'extraction d'information :

#### ● Difficulté de Question :

Il est calculé en fonction du score afin que la difficulté de la question (**difficile-moyen-facile**) puisse être déterminée, donc pour chaque question, on note :

Si la moyenne des scores de toutes les réponses est inférieure ou égale à 1,5, la question est considérée comme difficile, sinon, si la moyenne est supérieure à 1,5 et inférieure à 3,5, la question est moyenne, sinon la question est facile.

### 4.2.5 Processus de l'apprentissage automatique :

Le processus suivis dans la partie de l'apprentissage automatique de notre travail passe par l'étape de partitionnement de données, extraction de caractéristiques, les modèles utilisées et l'évaluation des modèles.

- a) L'étape de « partitionnement de données » consiste à diviser nos ensembles de données en 80% données d'entraînements et 20% données de test
- b) La seconde étape est l'étape de « extraction des caractéristiques » pour que nous puissions par la suite sélectionner la meilleure combinaison pour nos modèles guidés par l'entraînement.

Afin d'atteindre le modèle final ayant le meilleur résultat, nous mettons en oeuvre de nombreux algorithmes, parmi ces modèles nous avons testé : LR, SVR, DTR, Mlpregressor(c'est le modèle ANN) de la bibliothèque «Scikit-learn»( <https://scikit-learn.org/>) afin de prédire une note automatiquement en utilisant la propriété «Average\_Mark » où nous en dépendions comme sortie de notre modèle dans la partie entraînement La dernière étape est destinée non seulement pour la validation du modèle (pour avoir le modèle le plus performant) mais aussi pour le choix de la combinaison des caractéristiques qui nous mène à identifier le modèle final. a) Le but de notre évaluation est d'avoir une précision élevée avec une petite marge d'erreur cette dernière détermine la différence entre les notes prédites et les notes manuelles données par les enseignants. Pour le faire plusieurs

métriques d'évaluations sont utilisées nous citons : RMSE, Recall, Accuarcy, Corrélacion de Pearson.

#### 4.2.6 Environnement du développement :

En raison du volume de travail fourni et des ressources utilisées, le développement avec nos machines personnelles était pour nous comme un obstacle, car la RAM de 4 Go ne supporte pas le gigantesque de l'espace sémantique de 30 000 mots et le Word embeddings de 6 millions mots, ces difficultés ont été relativement surmontées, à l'aide du serveur fourni par **Google**, appelé **Colab**, qui fournit 35 Go du RAM.

### 4.3 Résultat et discussion :

Après un certain nombre de tentatives et d'ajustement pour le choix du modèle et des caractéristiques, le modèle final a été défini ainsi que les propriétés qui donnent le meilleur résultat.

Nous entamons la section des résultats obtenus en appliquant l'apprentissage automatique, où nous montrons les modèles utilisés, le meilleur-modèle, les caractéristiques ayant les plus d'impacts sur les résultats et une analyse approfondie du meilleur model obtenu.

Le tableau suivant présente les résultats de ces caractéristiques sur les différents algorithmes d'apprentissage automatique sur le « Dataset Arabe »

Types de modèle	Scores Entraînement	Scores Test	Corrélation	RMSE
LR	59,25%	60,18%	77,95%	0,8967
SVR				
DTR				
ANN				

FIGURE 4.8 – Tableau 4.3 : resultat

Enfin, comme perspectives, Il est evident que ce projet n'est pas une œuvre parfaite mais nous souhaitons et esperons presenter d'autres resultats que nous pouvons tirer de diverses analyses de l'expérience. Aussi, a étendre la modelisation proposée a différentes

approches de l'apprentissage automatique pour l'évaluation automatique des réponses courtes en langue arabe

# Conclusion générale

Le domaine de l'évaluation automatisée est un vaste domaine qui nécessite plusieurs ressources et de multiples techniques avec une bonne compréhension et une bonne ingénierie.

La difficulté de traiter la langue, en particulier les langues qui se distinguent par la richesse de leur banque de langues, comme la langue arabe, nous nous sommes donc concentrés dans cette étude principalement sur la langue arabe, en essayant d'enrichir les ouvrages. À cet égard, nous avons veillé à profiter du développement important du machine learning et de son intégration dans l'éducation pour l'explorer dans notre travail.

# Bibliographie

- [1] Mohler, M., Mihalcea, R. (2009). Text-to-text semantic similarity for automatic short answer grading. In Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (pp. 567-575). Association for Computational Linguistics.
- [2] Bloom taxonomie : A taxonomy for learning, teaching, and assessing : a revision of Bloom's taxonomy of educational objectives. New York : Longman,c2001
- [3] S. Burrows, I. Gurevych, and B. Stein, The Eras and Trends of Automatic Short Answer Grading, vol. 25.2015.
- [4] Singh, P. (1997). Review Essay : Basil Bernstein (1996). Pedagogy, symbolic control and identity. London : Taylor Francis. In British Journal of Sociology of Education,18(1),pp.119-124.
- [5] Leacock, C. and Chodorow, M. (2003). C-rater : Automated Scoring of Short-Answer Questions. Computers and the Humanities,37(4),389–405.
- [6] Dezso Sima, Balazs Schmuck, Sandor Szollosi, Arpad Miklos. (2007). Intelligent short text assessment in eMax. AFRICON 2007.
- Jordan, S.andMitchell, T.(2009).e-AssessmentforLearning ?ThePotentialofShort- AnswerFree-TextQuestions with Tailored Feedback. British Journal of Educational Technology,40(2),371–385.
- [8] Bukai, O., Pokorny, R., and Haynes, J. (2006). An Automated Short-Free-Text Scoring System : Development and Assessment. In Proceedings of the Twentieth Interser-

vice/Industry Training, Simulation, and Education Conference, pages 1–11. National Training and Simulation Association.

[9] Mohler, Michael Mihalcea, Rada (2004) Text-to-text Semantic Similarity for Automatic Short Answer Grading.

[10] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H. (2009). The WEKA Data Mining Software : An Update. SIGKDD Explorations, 11(1), 10 –18.

[11] Bailey, S. and Meurers, D. (2008). Diagnosing Meaning Errors in Short Answers to Reading Comprehension Questions. In J. Tetreault, J. Burstein, and R. De Felice, editors, Proceedings of the Third ACL Workshop on Innovative Use of NLP for Building Educational Apps.

[12] Madnani, N., Burstein, J., Sabatini, J., and Reilly, T. O. (2013). Automated Scoring of a Summary Writing Task Designed to Measure Reading Comprehension. In J. Tetreault, J. Burstein, and C. Leacock, editors, Proceedings of the Eighth Workshop on Innovative.

[13] [lig-membres.imag.fr/wang/Corpus/These.html/html/ch2](http://lig-membres.imag.fr/wang/Corpus/These.html/html/ch2)

[14] P. Resnik, *Parallel Strands : A Preliminary Investigation into Mining the Web for Bilingual Text, in Conference of the Association for Machine Translation in the Americas*, 1998.

[15] F. Issac, T. Hamon, L. Bouchard, L. Emirkanian, C. Fouqueré, *Extraction in Formatique de Données sur le Web : une expérience, in Multimédia, Internet et Francophonie : à la recherche d'un dialogue, Vancouver, Canada, mars 2001*

[16] GHOUL, Dhaou. *Construction d'un corpus arabe à partir du Web dans le but d'identifier les mots – outils ou tokens. JADT'14, Paris, France, 2 – 6 juin 2014.*

[17] ATOUBYasmine, BENAYADA Sma « Mesures de similarité sémantique pour un système d'évaluation automatique des réponses courtes : Application à la langue arabe » .Mémoire master USDB1.2018/2019.

[18] D.L.T.Rohde, L.M.Gonnerman, et D.C.Plaut, « An Improved Method for Deriving Word Meaning from Lexical », *Cogn. Psychol.*, vol.7, p.573-605, 2004.

[19] W.H.Gomaa et A.A.Fahmy, « A Survey of Text Similarity Approaches », *Int.J.Comput.Appl.*, vol.68, no13, p.13-18, 2013.

[20] R.Mihalea, C.Cole, et C.\tappaaa, *Copus – based and Knowledge – based Measures of Text Similarity*, pp.775–780, 2005.

D.Karani, « Introduction to Word Embedding and Word2Vec » by Dhruvil Karani | Towards Data Science, 01 – Sep – 2018. [Online]. Available : <https://towardsdatascience.com/introduction-to-word-embedding-and-word2vec-652d0c2060fa>. [Accessed : 25 – Aug – 2020].

[22] M.A.Zahran, A.Magooda, A.Y.Mahgoub, H.Raafat, M.Rashwan, and A.Atyia, « Word representations in vector space and their applications for Arabic », in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2015, vol.9041, pp.430–443, doi : 10.1007/978-3-319-18111-0\_32.

[23] « NLP word embeddings repository. » [Online]. Available : <http://vectors.nlpl.eu/repository/?fbclid=IwAR3xfVYqEVYovgrcW5GwWL57nSeLmUBzMV4EUNYD61xkXUYrqi7aT08>. [Accessed : 25 – Aug – 2020].

[24] M.Naili, A.H.Chaibi, dan H.H.BenGhezala, « Comparative Study of Word Embedding Methods in Text Mining ».

[25] K. Abainia, S. Ouamour, and H. Sayoud, “A novel robust Arabic light stemmer,” *J. Exp. Theor. Artif. Intell.*, vol. 29, no. 3, pp. 557–573, 2017.

[26] M. Mustafa, A. S. Eldeen, S. Bani-Ahmad, and A. O. Elfaki, “A Comparative Survey on Arabic Stemming : Approches and Challenges,” *Intell. Inf. Manag.*, vol. 9, no. 02, p. 39, 2017.

[27] T. Zerrouki, “Tashaphyne, Arabic light Stemmer/segment.” 2010.

## ملخص

تقييم المتعلم هو الحلقة الأضعف في منصات التدريب الجديدة عبر الإنترنت اليوم باستثناء أسئلة الاختيار من متعدد، بحيث لا تزال مهمة يدوية أو يميل إلى أن يكون ثنائيًا (صحيحًا أو خاطئًا) في هذا العمل نحن مهتمون بالأسئلة ذات الإجابات المختصرة وتقييمهم بواسطة التعلم الآلي .  
كلمات مفتاحية : التقييم الآلي للإجابات القصيرة، إدراج الكلمات، الفضاء الدلالي، التعلم الآلي

## Abstract

Learner assessment is the weakest link in today's new online training platforms with the exception of multiple choice questions, learner assessment is still a manual task or tends to be binary (true or false) in this task, We are interested in short answer questions and their machine learning assessment.

**Key words :** Automatic evaluation of short answers, word embedding, semantic space, Machine Learnin . . .

## Résumé

L'évaluation des apprenants est le maillon faible des plateformes de formation en ligne d'aujourd'hui arrivée a l'exception des questions à choix multiples, l'évaluation de l'apprenant c'est toujours une tâche manuelle ou a tendance à être binaire (vrai ou faux) dans cette tâche, Nous sommes intéressés par les questions à réponse courte et leur évaluation de l'apprentissage automatique.

**Mots clés :** Évaluation automatique des réponses courtes, Incorporation des mots, espace sémantique, Apprentissage Automatique . . .