

Sujets de Projets tutorés

Introduction

Un projet tutoré en analyse de données est une activité pédagogique au cours de laquelle les étudiants mettent en application leurs connaissances et compétences acquises en matière de collecte, traitement et analyse des données. Ce type de projet, encadré par un enseignant ou un tuteur, vise à développer à la fois des compétences techniques avancées en science des données et des compétences transversales comme la gestion de projet et la collaboration en équipe. Les projets tutorés en analyse de données offrent une opportunité unique de résoudre des problématiques réelles grâce à l'exploitation de données, tout en consolidant les concepts théoriques et les compétences pratiques. Ce projet est un tremplin pour les étudiants qui se destinent à une carrière dans le domaine de la science des données, de l'intelligence artificielle ou des systèmes décisionnels.

I. Contexte

Le projet tutoré s'adresse aux étudiants de troisième année en analyse de données et intelligence artificielle, qui se préparent à intégrer le marché du travail ou à effectuer un stage. Ce projet permet de consolider les compétences acquises durant les années de formation tout en confrontant les étudiants à des problématiques concrètes d'analyse et de traitement de données. Les étudiants devront traiter des jeux de données réels pour en extraire des informations utiles, construire des modèles prédictifs ou descriptifs, et présenter des résultats sous forme de tableaux de bord ou de rapports analytiques. L'objectif est de préparer les étudiants à manipuler des données dans un cadre professionnel et à maîtriser les outils, méthodes et bonnes pratiques liés à l'analyse de données.

II. Objectifs généraux

Les objectifs généraux du projet tutoré en analyse de données sont les suivants :

1. Appliquer les connaissances et compétences acquises en matière de collecte, prétraitement, analyse et visualisation des données.
2. Développer des compétences en modélisation de données, en intelligence artificielle, et en apprentissage automatique.
3. Se familiariser avec les outils et frameworks utilisés dans le domaine de l'analyse de données (Python, Pandas, Scikit-learn, etc.).
4. Acquérir des compétences transversales telles que la gestion de projet, la communication des résultats, et le travail en équipe.
5. Préparer les étudiants à leur future carrière dans le domaine de la science des données ou à des stages spécialisés.

III. Objectifs spécifiques

1. Collecter et préparer un jeu de données adapté au problème à résoudre.
2. Explorer et analyser les données pour en extraire des insights significatifs (analyse exploratoire).
3. Mettre en place des modèles d'analyse prédictive ou descriptive (modèles statistiques, machine learning).
4. Effectuer une validation rigoureuse des modèles et analyser leurs performances.
5. Visualiser les résultats et communiquer les analyses de manière compréhensible via des tableaux de bord et rapports.
6. Utiliser des outils de gestion de projet et de versionnage (Git, Trello) pour faciliter le travail en équipe et la collaboration.

IV. Résultats attendus

À la fin du projet tutoré, les résultats suivants sont attendus :

1. Une solution fonctionnelle d'analyse de données répondant à la problématique définie (modèles d'apprentissage automatique, rapports analytiques, etc.).
2. Un rapport complet incluant la description des données, les méthodes utilisées, les résultats obtenus et les conclusions tirées.
3. Une documentation détaillant l'architecture du projet, les méthodes employées, et les scripts utilisés pour la collecte, l'analyse et la modélisation des données.
4. Une présentation du projet mettant en évidence les conclusions tirées des analyses, les choix méthodologiques, et les recommandations basées sur les résultats.
5. Le développement des compétences techniques (manipulation de données, modélisation, visualisation) et transversales (travail en équipe, communication) chez les étudiants.

V. Description des projets

9. Chatbot pour la sensibilisation aux droits et lois

Description

Un chatbot conçu pour informer les citoyens sur leurs droits et sur les lois en vigueur au Burkina Faso. Il permet aux utilisateurs de poser des questions concernant leurs droits civils, les procédures légales, et fournit des informations sur les ressources juridiques disponibles.

Objectif Général

Créer un chatbot juridique capable d'informer les citoyens sur leurs droits, les procédures judiciaires et les lois nationales.

Fonctionnalités Clés

- Réponses aux questions courantes sur les droits civils, le droit du travail, et les lois foncières.
- Accès à des informations sur les démarches administratives et les services juridiques locaux.
- Suivi personnalisé des cas juridiques simples (ex. dépôt de plainte, obtention de documents).
- Notifications pour informer des évolutions législatives.

Technologies

- **Backend:** Flask (Python)
- **NLP:** NLTK ou SpaCy pour interpréter les questions juridiques.
- **Base de données:** PostgreSQL pour stocker les informations légales et les requêtes des utilisateurs.

Critères de Succès

- Niveau d'adoption par les citoyens cherchant des conseils juridiques.
- Précision des réponses aux questions juridiques.
- Taux de satisfaction mesuré par le nombre de cas résolus via le chatbot.

VI. Cahiers de charges

9. Chatbot pour la sensibilisation aux droits et lois

Objectif Général

Développer un chatbot juridique pour informer les citoyens de leurs droits et des lois en vigueur au Burkina Faso.

Fonctionnalités Clés

- Réponses automatiques aux questions courantes sur le droit civil, le droit du travail et les procédures légales.
- Informations sur les démarches administratives et juridiques.
- Notifications des évolutions législatives et des réformes.
- Accès aux contacts d'avocats et de services juridiques locaux.

Technologies

- **Backend:** Flask (Python)
- **NLP:** SpaCy pour analyser les questions légales.
- **Base de données:** PostgreSQL pour stocker les lois et informations légales.

Critères de Succès

- Pertinence et précision des informations juridiques.
- Taux d'adoption par les citoyens.
- Impact sur l'accès aux services juridiques.

VII. Canevas de rédaction du rapport du projet tutoré

1. Introduction

- **Contexte du projet**
 - o Présentation générale du domaine (ex : data science, IA, etc.).
 - o Problématique à résoudre avec l'analyse de données.
- **Objectifs du projet**
 - o Identification des objectifs à atteindre à travers l'analyse de données.
 - o Impact attendu de l'analyse sur la prise de décision ou les processus à améliorer.
- **Description de la plateforme ou des outils utilisés**
 - o Présentation des outils ou plateformes de collecte et traitement des données.

2. Analyse des besoins

- **Description des données collectées**
 - o Origine des données (source de données, scraping, bases externes).
 - o Volume et type de données (structurées, non structurées, big data, etc.).
- **Analyse des utilisateurs cibles et de leurs besoins**
 - o Identification des bénéficiaires de l'analyse de données (clients, entreprises, secteur public).
 - o Besoins spécifiques des utilisateurs en termes d'insights ou de prédictions.
- **Étude de marché et concurrence**
 - o Analyse des solutions existantes utilisant des données similaires.
 - o Positionnement du projet par rapport aux approches existantes.

3. Préparation et Exploration des Données

- **Pré-traitement des données**
 - o Nettoyage des données (suppression des valeurs manquantes, correction des anomalies, etc.).
 - o Transformation des données (normalisation, agrégation, etc.).
- **Exploration des données**
 - o Analyse exploratoire des données (distribution, corrélations, outliers, etc.).
 - o Visualisation initiale des données (graphiques, tableaux).
- **Feature Engineering**
 - o Création de nouvelles variables pertinentes pour l'analyse ou les modèles.
 - o Sélection des features les plus importantes pour l'analyse.

4. Conception et Modélisation

- **Choix des techniques d'analyse et des modèles**
 - o Justification des méthodes statistiques, algorithmes d'apprentissage machine ou techniques de deep learning choisies.
 - o Présentation des alternatives et des choix finaux.
- **Description de l'architecture du système d'analyse**
 - o Schéma des processus d'analyse de bout en bout (collecte, traitement, modélisation, visualisation).
 - o Modélisation des données (diagramme des flux de données, etc.).
- **Modélisation de la base de données (si applicable)**
 - o Schéma de la base de données utilisée pour stocker les résultats de l'analyse.

5. Implémentation des modèles d'analyse

- **Explication des étapes de développement**
 - o Présentation des différentes étapes de la mise en place des modèles et du pipeline d'analyse.
- **Présentation des modèles et algorithmes développés**
 - o Détails sur les modèles implémentés (algorithmes de machine learning, IA).
 - o Explication des paramètres et hyperparamètres utilisés.
- **Description des modules et composants clés**

- o Explication des modules ou scripts responsables des étapes majeures (pré- traitement, prédiction, visualisation).

6. Résultats de l'analyse

- **Résultats des modèles et interprétation**
 - o Performance des modèles d'analyse (précision, rappel, F1-score, etc.).
 - o Analyse des résultats obtenus par rapport aux objectifs fixés.
- **Visualisation des résultats**
 - o Graphiques, tableaux, et autres moyens de présentation des résultats.
 - o Comparaison des résultats des différents modèles.

7. Tests et Validation

- **Description des méthodes de validation**
 - o Validation croisée, tests A/B, division en jeux de données d'entraînement et de test.
- **Résultats des tests et analyse des erreurs**
 - o Analyse des erreurs et limites des modèles (erreurs de prédiction, biais, variance).
 - o Propositions d'amélioration basées sur les résultats.

8. Performances et Optimisation

- **Évaluation des performances des modèles**
 - o Temps de calcul, efficacité des algorithmes sur de grands ensembles de données.
- **Optimisation des modèles**
 - o Ajustements des hyperparamètres pour améliorer les performances.
 - o Techniques d'optimisation utilisées (pruning, grid search, etc.).

9. Sécurité et protection des données

- **Mesures de sécurité des données**
 - o Explication des techniques de protection des données (cryptage, anonymisation).
 - o Respect des normes de confidentialité (RGPD, protection des données sensibles).
- **Gestion des données personnelles**

- o Politique de stockage et de traitement des données personnelles, si applicable.

10. Conclusion

- **Résumé des réalisations du projet**
 - o Synthèse des résultats obtenus et des objectifs atteints.
- **Perspectives d'amélioration**
 - o Pistes d'amélioration des modèles et techniques d'analyse utilisées.
- **Développement futur**
 - o Opportunités d'évolution et d'intégration à plus grande échelle.

11. Annexes

- **Diagrammes et visualisations complémentaires**
 - o Graphiques supplémentaires issus de l'analyse de données.
- **Extraits de code source**
 - o Code clé utilisé pour le traitement des données et les modèles.
- **Documentation des algorithmes et modèles**
 - o Documentation technique complémentaire (modèles statistiques, architecture, etc.).