

REPUBLIQUE DEMOCRATIQUE DU CONGO
ENSEIGNEMENT SUPERIEUR ET UNIVERSITAIRE
INSTITUT SUPERIEUR D'INFORMATIQUE ET DE GESTION



ISIG-GOMA
MASTER COMPLEMENTAIRE EN INFORMATIQUE APPLIQUEE A
LA GESTION DES ENTREPRISES

**Examen-projet : Développement d'une Application Web de Prédiction du Taux
de Conversion d'un Site Web à l'aide du Machine Learning**

TRAVAIL PRATIQUE DU COURS DE DATA MINING AND MACHINE LEARNING

Présenté par :

- 1. BURAKALI BAYONGWA Justin**
- 2. ZIHALIRWA NSHOMBO Alphonse**
- 3. PAUL MUPENDA**
- 4. HABAMUNGU DUNIA**

Dispensé par : Prof. Dr. Alain AKWIR

Promotion : M2P MIAGE-IMS

Année académique 2024-2025

1. Dataset :

<https://www.kaggle.com/datasets/anthonytherrien/website-traffic>

2. Lien du projet sur GitHub :

<https://github.com/JustinBB1/Website-Traffic-and-User-Engagement-Metrics.git>

3. Objectif du projet :

Concevoir une application web développée avec le framework Django, intégrant un modèle de Machine Learning, dans le but de, d' :

- Prédire le taux de conversion d'un site web à partir de ses principales métriques de performance issues du dataset Website Traffic and User Engagement Metrics.
- Analyser l'impact des différentes variables (telles que les pages vues, la durée moyenne des sessions, le taux de rebond, les clics, etc.) sur le comportement des utilisateurs et sur la conversion.
- Fournir un outil interactif et accessible permettant d'effectuer des prédictions en temps réel via un formulaire web intuitif.
- Evaluer les performances du modèle (MAE, MSE, RMSE, R^2) afin de garantir la fiabilité des prédictions intégrées dans l'application.
- Faciliter la prise de décision pour l'optimisation des performances web en exploitant les résultats générés par le modèle.

4. Structure du projet

website_traffic_predictor/

```

|— manage.py
|— website_traffic/
|   |— __init__.py
|   |— settings.py
|   |— urls.py
|   |— wsgi.py
|— traffic_app/
|   |— migrations/
|   |— __init__.py
|   |— admin.py
|   |— apps.py
|   |— models.py
|   |— views.py
|   |— urls.py
|   |— templates/
|       |— index.html
|       |— predict.html
|       |— results.html
|       |— recommend.html
|— static/
|   |— css/
|   |— js/

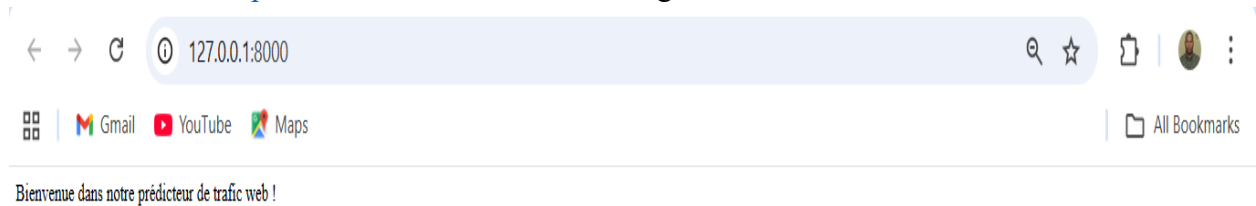
```

```
|  └─ images/  
|  └─ requirements.txt  
|  └─ README.md
```

5. Test de lancement du serveur :

`python manage.py runserver`

ouvrir l'adresse <http://127.0.0.1:8000/> dans le navigateur



6. Exploration des données explore_data.py :

```
C:\Windows\System32\cmd.exe
C:\Users\JBB\Documents\TP Machine Learning\Website Traffic and User Engagement Metrics\website_traffic>python scripts/explore_data.py

=====
ANALYSE DU DATASET : WEBSITE TRAFFIC
=====
Dataset chargé avec succès!

TAILLE DU DATASET : 2000 lignes, 7 colonnes

COLONNES DISPONIBLES :
1. Page Views (int64) - Exemple: 5
2. Session Duration (float64) - Exemple: 11.051381236591851
3. Bounce Rate (float64) - Exemple: 0.230651936613825
4. Traffic Source (object) - Exemple: Organic
5. Time on Page (float64) - Exemple: 3.890460070474997
6. Previous Visits (int64) - Exemple: 3
7. Conversion Rate (float64) - Exemple: 1.0

APERÇU DES DONNÉES (5 premières lignes) :
Page Views Session Duration Bounce Rate Traffic Source Time on Page Previous Visits Conversion Rate
0 5 11.051381 0.230652 Organic 3.890460 3 1.0
1 4 3.429316 0.391001 Social 8.478174 0 1.0
2 4 1.621052 0.397986 Organic 9.636170 2 1.0
3 5 3.629279 0.180458 Organic 2.071925 3 1.0
4 5 4.235843 0.291541 Paid 1.960654 5 1.0

STATISTIQUES DESCRIPTIVES :
Page Views Session Duration Bounce Rate Time on Page Previous Visits Conversion Rate
count 2000.000000 2000.000000 2000.000000 2000.000000 2000.000000 2000.000000
mean 4.950500 3.022045 0.284767 4.027439 1.978500 0.982065
std 2.183903 3.104518 0.159781 2.887422 1.432852 0.065680
min 0.000000 0.003613 0.007868 0.068515 0.000000 0.343665
25% 3.000000 0.815828 0.161986 1.935037 1.000000 1.000000
50% 5.000000 1.993983 0.266375 3.315316 2.000000 1.000000
75% 6.000000 4.197569 0.388551 5.414627 3.000000 1.000000
max 14.000000 20.290516 0.844939 24.796182 9.000000 1.000000

VALEURS MANQUANTES :
Page Views: Aucune valeur manquante
Session Duration: Aucune valeur manquante
Bounce Rate: Aucune valeur manquante
Traffic Source: Aucune valeur manquante
Time on Page: Aucune valeur manquante
Previous Visits: Aucune valeur manquante
Conversion Rate: Aucune valeur manquante

VALEURS UNIQUES PAR COLONNE :
Page Views: 15 valeurs uniques
Session Duration: 2000 valeurs uniques
Bounce Rate: 2000 valeurs uniques
Traffic Source: 5 valeurs -> ['Organic' 'Social' 'Paid' 'Direct' 'Referral']
Time on Page: 2000 valeurs uniques
Previous Visits: 10 valeurs uniques
Conversion Rate: 228 valeurs uniques

TRAFFIC SOURCE - DISTRIBUTION :
Traffic Source
Organic 786
Paid 428
Referral 301
Social 260
Direct 216
Name: count, dtype: int64

MATRICE DE CORRÉLATION (colonnes numériques) :
Page Views Session Duration Bounce Rate Time on Page Previous Visits Conversion Rate
Page Views 1.000000 -0.012788 0.030223 0.023941 0.028595 0.126635
Session Duration -0.012788 1.000000 -0.015931 -0.013985 -0.025534 0.177798
Bounce Rate 0.030223 -0.015931 1.000000 0.039340 -0.015860 -0.049051
Time on Page 0.023941 -0.013985 0.039340 1.000000 -0.028729 0.229669
Previous Visits 0.028595 -0.025534 -0.015860 -0.028729 1.000000 0.109496
Conversion Rate 0.126635 0.177798 -0.049051 0.229669 0.109496 1.000000

Heatmap sauvegardée : 'dataset/correlation_heatmap.png'

Échantillon sauvegardé : 'dataset/sample_data.csv' (50 premières lignes)
```

7. Entraînement : Exécuter 'train_model.py' :

```

C:\Windows\System32\cmd.exe
C:\Users\JBB\Documents\TP Machine Learning\Website Traffic and User Engagement Metrics\website_traffic>python scripts/train_model.py

=====
ENTRAÎNEMENT DU MODÈLE DE PRÉDICTION
=====
[+] Données chargées : 2000 lignes

[+] Variable cible : Conversion Rate
[+] Variables d'entrée : ['Page Views', 'Session Duration', 'Bounce Rate', 'Traffic Source', 'Time on Page', 'Previous Visits']

[+] Encodage des variables catégorielles...
[+] Encodeur sauvegardé : 'models/encoders/traffic_source_encoder.pkl'

[+] Dimensions :
    X (features) : (2000, 6)
    y (target)  : (2000,)

[+] Division train/test :
    Train : 1600 échantillons
    Test  : 400 échantillons

[+] Entraînement du modèle RandomForest...
[+] Modèle entraîné !

[+] ÉVALUATION DU MODÈLE :
    MAE : 0.0219
    MSE : 0.0025
    RMSE : 0.0497
    R² : 0.2104

[+] IMPORTANCE DES VARIABLES :
      feature  importance
    Time on Page  0.306346
    Session Duration 0.292722
    Bounce Rate   0.168628

    Page Views    0.121712
    Previous Visits 0.062672
    Traffic Source_encoded 0.047920

[+] Modèle sauvegardé : 'models/traffic_model.pkl'
[+] Métadonnées sauvegardées : 'models/model_metadata.json'

=====
ENTRAÎNEMENT TERMINÉ !
=====

[+] Prochaine étape : Intégrer le modèle dans Django

```

8. Intégration du modèle dans Django

8.1. Page d'accueil : <http://127.0.0.1:8000/>

The screenshot displays the 'WebTraffic Predictor' web application running on a local server at 127.0.0.1:8000. The browser's address bar shows the URL, and the page title is 'WebTraffic Predictor'. The application features a clean, modern design with a light blue and white color scheme.

Header: The top navigation bar includes the application logo, the name 'WebTraffic Predictor', and three menu items: 'Accueil', 'Prédiction', and 'À propos'.

Main Content Area:

- Prédicteur de Trafic Web:** The main heading is accompanied by a subtext: 'Utilisez l'intelligence artificielle pour prédire le taux de conversion de votre site web en fonction de ses métriques de performance.'
- Dataset Utilisé:** A section titled 'Dataset Utilisé' explains that the model is trained on the 'Website Traffic and User Engagement Metrics' dataset, which contains 2000 records with 7 characteristics:
 - Page Views:** Nombre de pages vues
 - Session Duration:** Durée de session (minutes)
 - Bounce Rate:** Taux de rebond (%)
 - Traffic Source:** Source du trafic
 - Time on Page:** Temps sur page (minutes)
 - Previous Visits:** Visites précédentes
 - Conversion Rate:** Taux de conversion (%)
- Buttons:** Below the dataset information, there are two buttons: 'Voir sur Kaggle' and 'Essayer la prédiction'.
- Fonctionnalités Principales:** A section titled 'Fonctionnalités Principales' highlights three key features:
 - Modèle IA Prédicatif:** Describes the Random Forest algorithm trained on 2000 samples for precise predictions.
 - Analyse en Temps Réel:** Describes the real-time prediction of conversion rates based on site metrics.
 - Recommandations Intelligentes:** Describes intelligent suggestions to improve site performance.
- Buttons:** Below the features, there are two identical blue buttons labeled 'Commencer Maintenant'.

Footer: The footer is divided into two columns. The left column contains the project title 'Projet Machine Learning', the goal 'Prédiction du taux de conversion des sites web', and the dataset/framework 'Dataset: Website Traffic | Framework: Django'. The right column lists the group 'Groupe 3 - ISIG-GOMA', the team members 'BURAKALI BAYONGWA Justin (Chef)', 'HABAMUNGU DUNIA | MUPENDA Paul | ZIHAIIRWA SHOMBO Alphonse', and the academic year 'Année académique 2024-2025'. A copyright notice '© 2024 WebTraffic Predictor. Tous droits réservés.' is centered at the bottom.

The Windows taskbar at the bottom shows the system clock as 18:29 on 03/12/2025, with a temperature of 22°C and various system icons.

8.2. Formulaire de prédiction et résultats : <http://127.0.0.1:8000/predict/>

The screenshot shows a web browser window with the URL <http://127.0.0.1:8000/predict/>. The page title is "WebTraffic Predictor". The main heading is "Prédire le Taux de Conversion". Below this, a note states: "Remplissez les champs ci-dessous avec les métriques de votre site web pour prédire le taux de conversion." The form contains six input fields arranged in two columns:

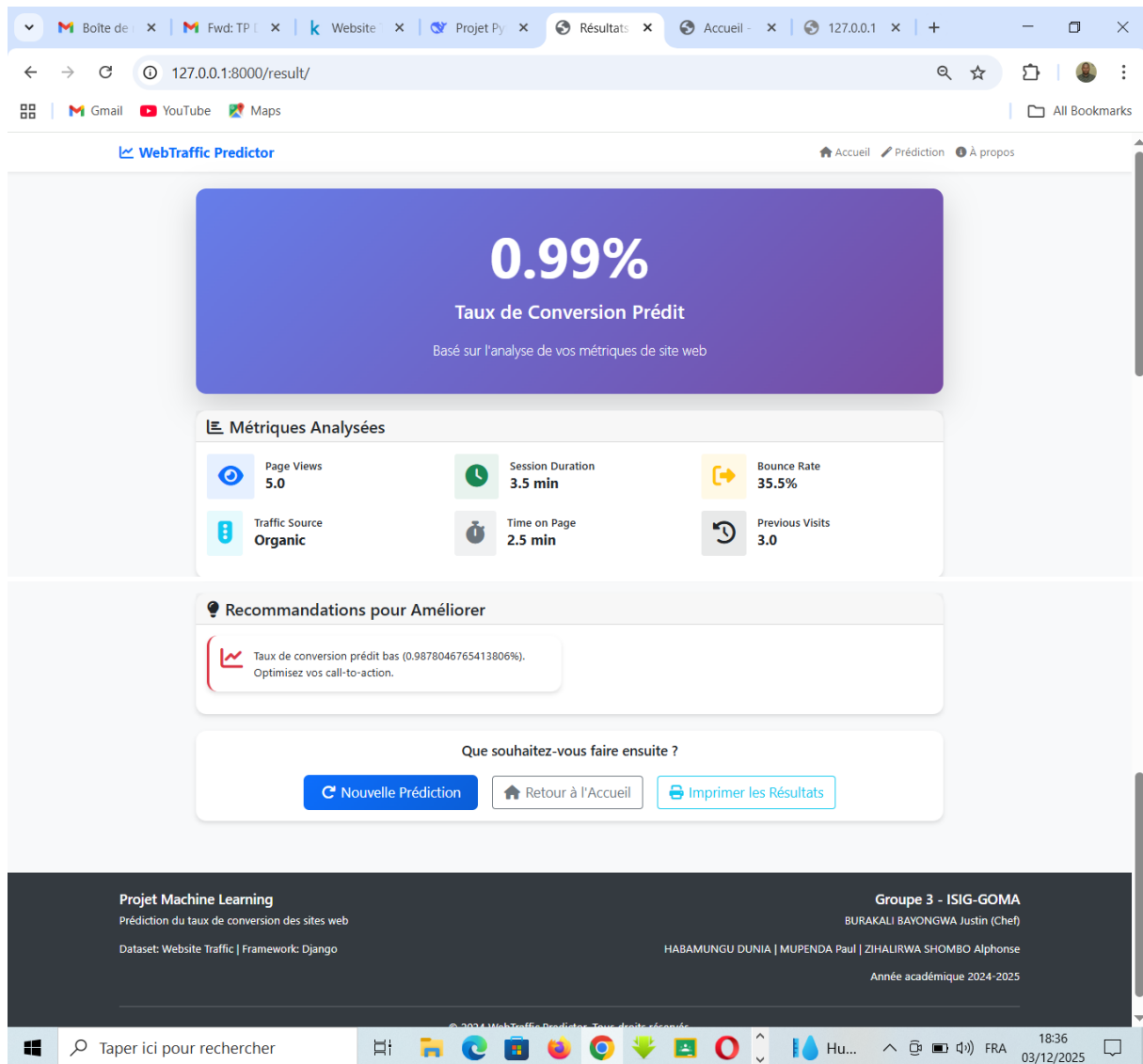
- Page Views:** Input field with value "5". Description: "Nombre de pages vues par session (1-50)".
- Session Duration (minutes):** Input field with value "3.5". Description: "Durée moyenne d'une session en minutes".
- Bounce Rate (%):** Input field with value "35.5". Description: "Pourcentage de visiteurs qui quittent après une page".
- Traffic Source:** Dropdown menu with "Organic" selected. Description: "Source principale du trafic".
- Time on Page (minutes):** Input field with value "2.5". Description: "Temps moyen passé sur une page".
- Previous Visits:** Input field with value "3". Description: "Nombre moyen de visites précédentes".

Below the form, a blue button labeled "Générer la Prédiction" and a white button labeled "Retour à l'accueil" are visible. A light blue note box states: "Note : Ces valeurs sont utilisées pour entraîner notre modèle Random Forest. La prédiction sera affichée sur la page suivante." Below this, a section titled "Conseils pour de meilleures prédictions" lists four tips:

- ✓ Utilisez des données réelles de Google Analytics si possible
- ✓ Pour le Bounce Rate, visez entre 20% et 40% pour un site standard
- ✓ La durée de session idéale varie entre 2 et 5 minutes
- ✓ Le trafic "Organic" (organique) a généralement un meilleur taux de conversion

The footer of the page contains the text "Projet Machine Learning" and "Prédiction du taux de conversion des sites web". On the right, it says "Groupe 3 - ISIG-GOMA" and "BURAKALI BAYONGWA Justin (Chef)". The bottom of the browser window shows the Windows taskbar with the search bar "Taper ici pour rechercher" and the system clock showing "18:34" on "03/12/2025".

Résultats :



Impression des résultats

The screenshot shows a web browser window with multiple tabs. The active tab is 'Résultats' at the URL '127.0.0.1:8000/result/'. The browser's address bar shows the URL and navigation icons. The page content is a 'WebTraffic Predictor' dashboard. It features a large card displaying '0.99%' as the 'Taux de Conversion Prédit' (Predicted Conversion Rate), with a subtitle 'Basé sur l'analyse de vos métriques de site web' (Based on the analysis of your website metrics). Below this, a section titled 'Métriques Analysées' (Analyzed Metrics) lists several key performance indicators: Page Views (5.0), Session Duration (3.5 min), Bounce Rate (35.5%), Traffic Source (Organic), Time on Page (2.5 min), and Previous Visits (3.0). At the bottom of the dashboard, there is a section for 'Recommandations pour Améliorer' (Recommendations for Improvement). A print overlay is visible on the right side of the browser window, titled 'Print' and indicating '2 sheets of paper'. The print settings include: Destination (HP LaserJet Professional), Pages (All), Copies (1), and Layout (Portrait). There are 'Print' and 'Cancel' buttons at the bottom of the print overlay. The Windows taskbar at the bottom shows the search bar with the text 'Taper ici pour rechercher' (Type here to search), several application icons, and the system clock showing 18:38 on 03/12/2025.

127.0.0.1:8000/result/

WebTraffic Predictor

0.99%

Taux de Conversion Prédit

Basé sur l'analyse de vos métriques de site web

Métriques Analysées

- Page Views: 5.0
- Session Duration: 3.5 min
- Bounce Rate: 35.5%
- Traffic Source: Organic
- Time on Page: 2.5 min
- Previous Visits: 3.0

Recommandations pour Améliorer

Print

2 sheets of paper

Destination: HP LaserJet Professional

Pages: All

Copies: 1

Layout: Portrait

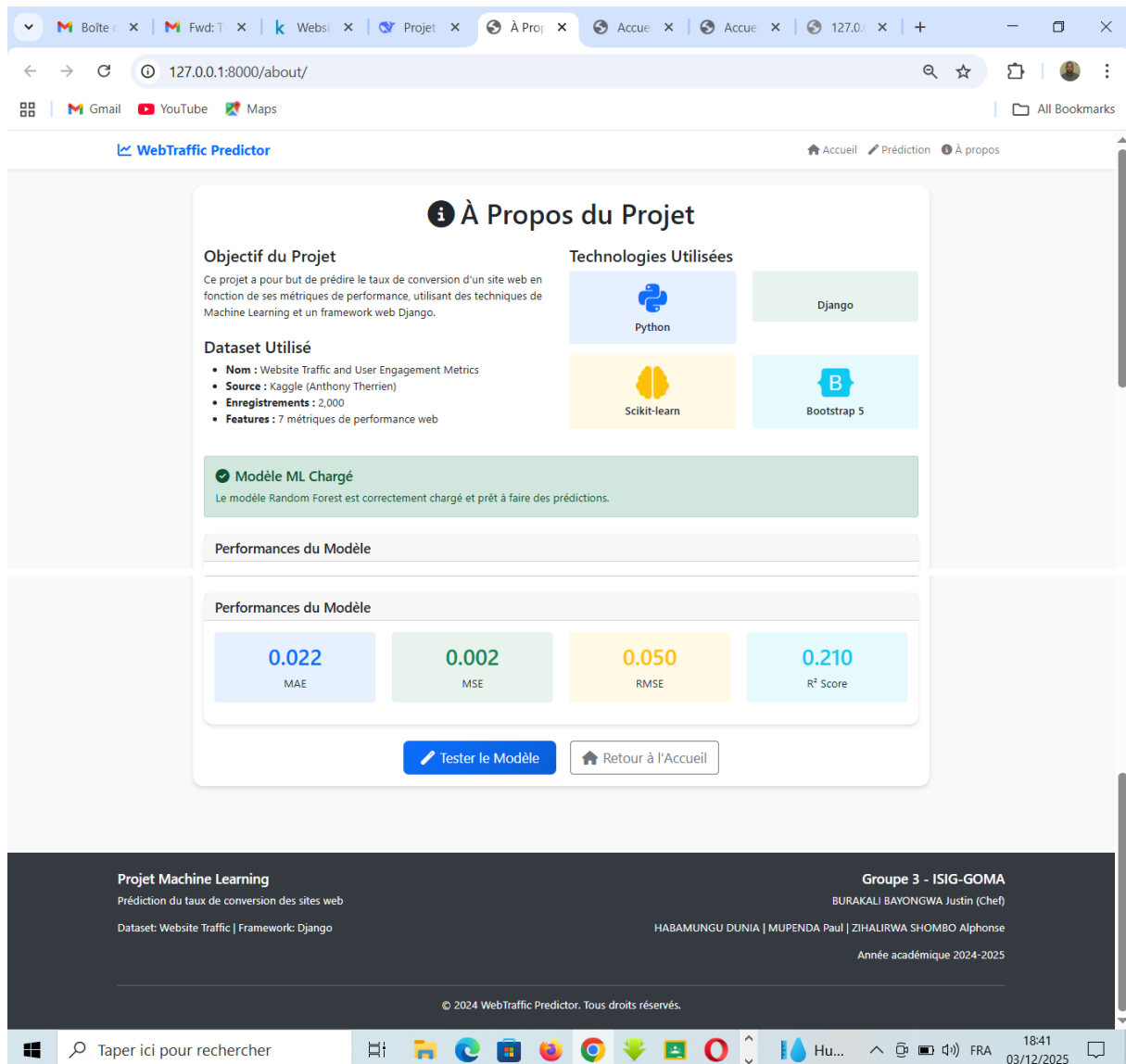
More settings

Print Cancel

Taper ici pour rechercher

18:38 03/12/2025

8.3. Page 'À propos/About' : <http://127.0.0.1:8000/about/>



Conclusion :

Au terme de ce projet, nous avons développé une application web fonctionnelle qui intègre un modèle de Machine Learning capable de prédire le taux de conversion d'un site web à partir de ses principales métriques de performance. En exploitant le dataset Website Traffic and User Engagement Metrics, nous avons pu analyser les facteurs influençant la conversion et construire un modèle pertinent, basé notamment sur l'algorithme Random Forest, dont les performances ont été évaluées à travers des indicateurs tels que le MAE, le MSE, le RMSE et le R² Score.

L'intégration du modèle dans une application Django a permis de mettre en place une interface simple, interactive et ergonomique, offrant aux utilisateurs la possibilité d'effectuer des prédictions en temps réel via un formulaire web. Cette approche démontre non seulement

la faisabilité technique du projet, mais aussi la valeur pratique de l'utilisation des techniques de data mining et de machine learning dans la prise de décision orientée données.

Ce TP contribue ainsi à illustrer comment les outils d'analyse prédictive peuvent aider les gestionnaires de sites web à mieux comprendre le comportement des utilisateurs, à optimiser leurs stratégies numériques et, au final, à améliorer les performances globales de leurs plateformes. Des perspectives d'amélioration sont possibles, notamment l'utilisation de modèles plus avancés, l'intégration de données supplémentaires ou la mise en place de visualisations plus poussées, afin de renforcer encore la précision et l'utilité de l'application.