

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/anthonysterrien/website-traffic>

2. Lien du projet sur GitHub :

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

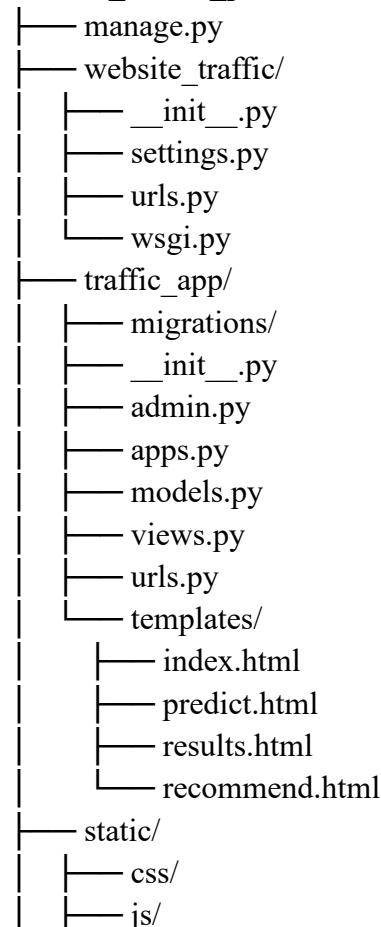
3. Objectif du projet :

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, R²) 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/



```
|   └── 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

C:\Windows\System32\cmd.exe
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     269
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

Heatmap sauvegardée : 'dataset/correlation_heatmap.png'

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

7. Entrainement : 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
=====
[D] Données chargées : 2000 lignes
[D] Variable cible : Conversion Rate
[D] Variables d'entrée : ['Page Views', 'Session Duration', 'Bounce Rate', 'Traffic Source', 'Time on Page', 'Previous Visits']
[D] Encodage des variables catégorielles...
[D] Encodeur sauvegardé : 'models/encoders/traffic_source_encoder.pkl'

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

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

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

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

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

=====
C:\Windows\System32\cmd.exe
      Page Views    0.121712
      Previous Visits  0.062672
  Traffic Source_encoded  0.047920

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

=====
ENTRAÎNEMENT TERMINÉ !
=====

[D] 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 shows the homepage of the "WebTraffic Predictor" application. At the top, there's a navigation bar with tabs like "Accueil - Pré", "127.0.0.1", and "À propos". Below the header, the title "Prédicteur de Trafic Web" is displayed with a subtitle: "Utilisez l'intelligence artificielle pour prédire le taux de conversion de votre site web en fonction de ses métriques de performance." A section titled "Dataset Utilisé" describes the training dataset: "Website Traffic and User Engagement Metrics" with 2000 records and 7 characteristics. It lists metrics: Page Views, Session Duration, Bounce Rate, Traffic Source, Time on Page, Previous Visits, and Conversion Rate. Buttons for "Voir sur Kaggle" and "Essayer la prédition" are present. The main content area is titled "Fonctionnalités Principales" and includes three cards: "Modèle IA Prédictif" (Random Forest algorithm), "Analyse en Temps Réel" (instantaneous conversion rate prediction), and "Recommandations Intelligentes" (smart performance suggestions). A large blue button "Commencer Maintenant" is centered below these cards. The footer contains project details: "Projet Machine Learning", "Prédiction du taux de conversion des sites web", "Dataset: Website Traffic | Framework: Django", "Groupe 3 - ISIG-GOMA", "BURAKALI BAYONGWA Justin (Chef)", "HABAMUNGU DUNIA | MUPENDA Paul | ZIHALIRWA SHOMBO Alphonse", "Année académique 2024-2025", and copyright information: "© 2024 WebTraffic Predictor. Tous droits réservés." The bottom bar also shows a search bar, system icons, and a date/time stamp: "18:29 03/12/2025".

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

Prédire le Taux de Conversion

Remplissez les champs ci-dessous avec les métriques de votre site web pour prédire le taux de conversion.

Page Views <input type="text" value="5"/>	Session Duration (minutes) <input type="text" value="3.5"/>
Nombre de pages vues par session (1-50)	Durée moyenne d'une session en minutes
Bounce Rate (%) <input type="text" value="35.5"/>	Traffic Source <input type="text" value="Organic"/>
Pourcentage de visiteurs qui quittent après une page	Source principale du trafic
Time on Page (minutes) <input type="text" value="2.5"/>	Previous Visits <input type="text" value="3"/>
Temps moyen passé sur une page	Nombre moyen de visites précédentes

Note : Ces valeurs sont utilisées pour entraîner notre modèle Random Forest. La prédiction sera affichée sur la page suivante.

Générer la Prédiction **Retour à l'accueil**

Conseils pour de meilleures prédictions

- ✓ 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

Projet Machine Learning
Prédiction du taux de conversion des sites web
Dataset: Website Traffic | Framework: Django

Groupe 3 - ISIG-GOMA
BURAKALI BAYONGWA Justin (Chef)
HABAMUNGU DUNIA LMURENDA Paul LIZAHIRWA SHOMBO Alphonse

1834 FRA 03/12/2025

Taper ici pour rechercher

Résultats :

The screenshot shows a web-based traffic predictor application. At the top, the URL is 127.0.0.1:8000/result/. The main header reads "WebTraffic Predictor". A large purple box displays the predicted conversion rate: **0.99%**. Below it, the text "Taux de Conversion Prédict" and "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

- Taux de conversion prédict bas (0.9878046765413806%). Optimisez vos call-to-action.

Que souhaitez-vous faire ensuite ?

[Nouvelle Prédiction](#) | [Retour à l'Accueil](#) | [Imprimer les Résultats](#)

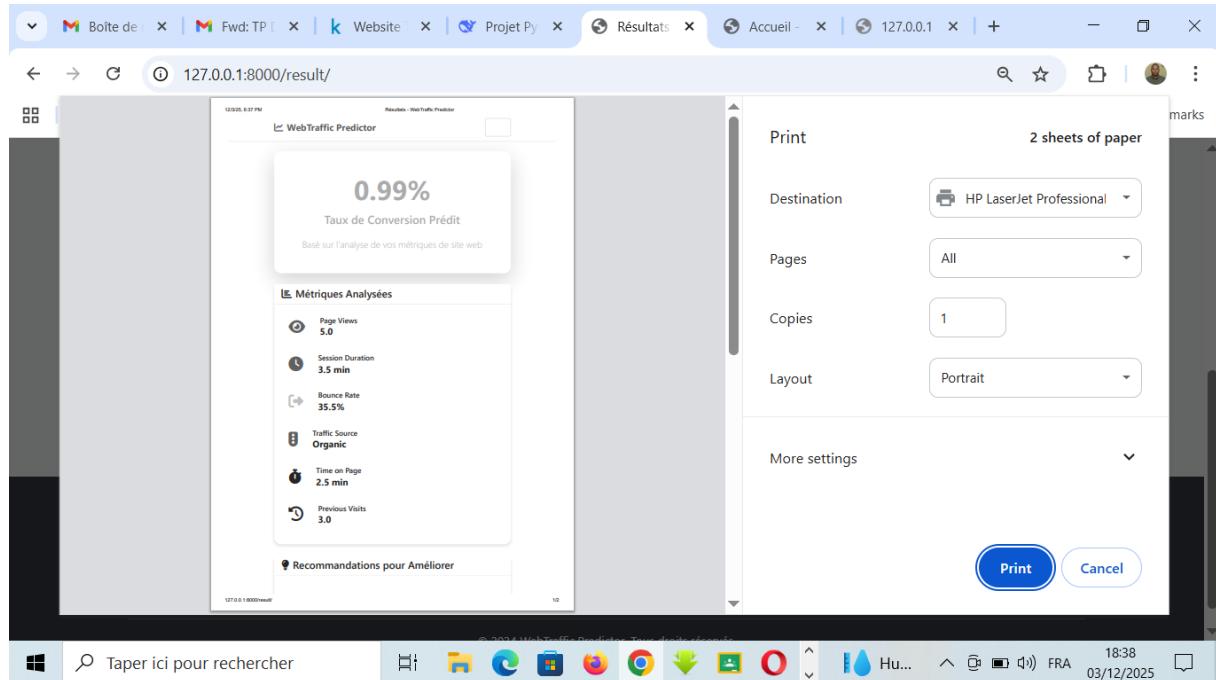
Projet Machine Learning
Prédiction du taux de conversion des sites web
Dataset: Website Traffic | Framework: Django

Groupe 3 - ISIG-GOMA
BURAKALI BAYONGWA Justin (Chef)
HABAMUNGU DUNIA | MUPENDA Paul | ZIHALIRWA SHOMBO Alphonse
Année académique 2024-2025

© 2024 WebTraffic Predictor. Tous droits réservés.

Windows taskbar at the bottom: Search bar, Start button, Task View, File Explorer, Edge, Taskbar settings, Volume, Battery, Language, Date/Time (18:36, 03/12/2025).

Impression des résultats



8.3. Page ‘A propos/About’ : <http://127.0.0.1:8000/about/>

The screenshot shows the 'About' page of a web application. At the top, there's a navigation bar with various tabs like 'Boîte', 'Fwd:T', 'Websi', 'Projet', 'À Proj...', 'Accue...', 'Accue...', '127.0...', and a '+' button. Below the navigation is a toolbar with icons for Gmail, YouTube, Maps, and 'All Bookmarks'. The main content area has a header 'À Propos du Projet' with a subtitle 'Objectif du Projet'. It describes the project's goal of predicting website conversion rates using Machine Learning and Django. A section 'Dataset Utilisé' lists the source as Kaggle (Anthony Therrien), with 2,000 records and 7 features. 'Technologies Utilisées' includes Python, Django, Scikit-learn, and Bootstrap 5. A green box indicates the 'Modèle ML Chargé' (Random Forest) is loaded. Below this are sections for 'Performances du Modèle' showing MAE (0.022), MSE (0.002), RMSE (0.050), and R² Score (0.210). Buttons for 'Tester le Modèle' and 'Retour à l'Accueil' are present. The footer contains project information: 'Projet Machine Learning', 'Prédiction du taux de conversion des sites web', 'Dataset: Website Traffic | Framework: Django', 'Groupe 3 - ISIG-GOMA', 'BURAKALI BAYONGWA Justin (Chef)', 'HABAMUNGU DUNIA | MUPENDA Paul | ZIHALIRWA SHOMBO Alphonse', 'Année académique 2024-2025', and copyright notice '© 2024 WebTraffic Predictor. Tous droits réservés.' The bottom bar shows a search bar, taskbar icons, and system status.

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.