



19-11-2021

Rapport d'avancement

Tentative d'apprentissage des corrélations entre phases de la lune et du soleil et coefficients des marées

Auteurs :

Otmane EL ALOI, Abdelhafid SAOUD, Abdelhadi ZIANE

Encadrant :

Mr. Olivier ROUX

Option :
Informatique

Table des matières

1	Feature engineering : Exploitation des variables liée à la physique des marées	2
1.1	Explotation uniquement des variables temporelles :	2
1.2	Explotation uniquement des variables de la physique :	3
1.3	Etude d'importance : Quelle variable permet d'expliquer plus de variance dans l'espace de donnée	4
1.4	Prédiction des coefficients en exploitant uniquement les grandeurs physiques	6
2	L'interface WEB	8
2.1	Home page	8
2.2	Page de prédiction	9
2.3	Co-founders page	10
3	Planning	11
3.1	Planning de la semaine en cours	11
3.2	Planning de la semaine prochaine	12

Avancement

Au cours de cette semaine, nous avons continué l'amélioration du modèle des forêts aléatoires et ceci en exploitant le pouvoir prédictive des variables temporelles et d'autres issue des lois physiques.

Cette approche s'avère meilleur que l'approche précédente qui s'appuyait uniquement sur les corrélations entre la valeur du coefficient de marée à un instant t et ses valeurs antérieurs.

Nous décrivons dans les sections qui suivent, les résultats des expériences effectuées jusqu'à présent.

1 Feature engineering : Exploitation des variables liée à la physique des marées

1.1 Explotation uniquement des variables temporelles :

Dans cette expérience nous avons employés uniquement la corrélation entre la variable à prédire (coefficient de marrée) et les variables temporelles associée au valeurs antérieurs. Nous avons aussi étudié les corrélations en exploitant deux types de calendrier :

- **Le calendrier Grégorien**
- **Le calendrier Hijri** : L'idée derrière l'utilisation de ce calendrier et qu'un mois est déterminé par une tour complète de la lune sur la terre. Ainsi la corrélation entre marrée et les dates du calendrier hijri serait plus fortes que les corrélations entre les dates du calendrier grégorien.

Ci-dessous le résultat de prédiction en exploitant uniuqent les variables temporelles :

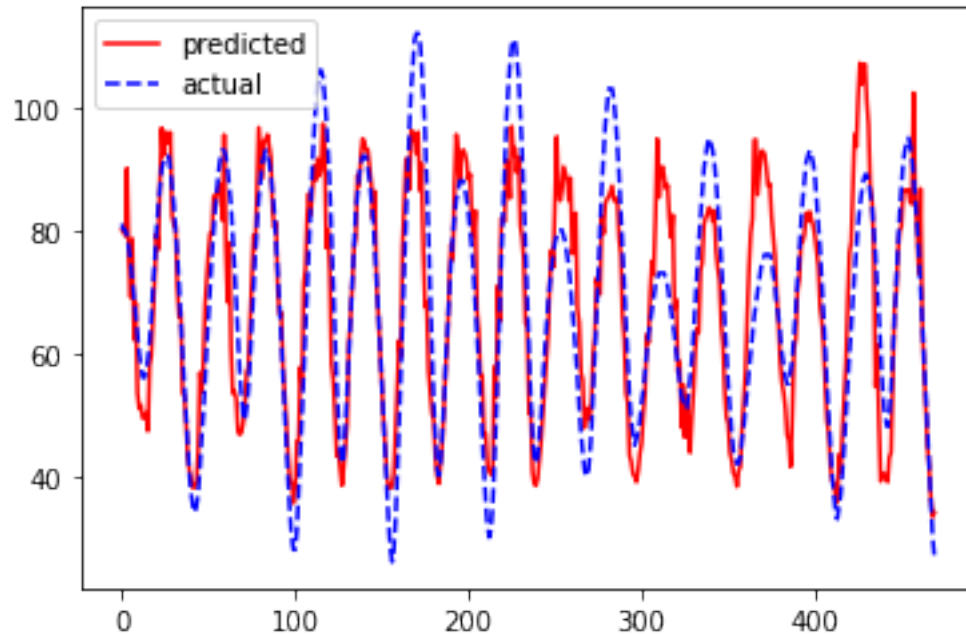


FIGURE 1 – Résultat de prédiction des coefficients de marrée de l'année 2021 avec les variables temporelles.

Le modèle utilisé pour faire la prédiction est toujours le modèle des forêts aléatoires et la métrique d'évaluation est la racine des moyennes quadratiques.

$$RMSE(coef_predicted, coef_actual) = 9.95$$

1.2 Explotation uniquement des variables de la physique :

Nous avons exploité une bibliothèque python qui permet à partir d'une date donnée de fournir des grandeurs physiques associées aux astres du système solaire, les grandeurs extraites sont comme suit :

- moon_phase
- earth_moon_distance
- sun_moon_distance
- libration_lat
- libration_long
- subsolar_latitude
- elongation
- earth_sun_distance

D'après la figure qui suit, les grandeurs physique décrit précédemment semblent fortement corrélées avec notre variable cible (coefficient de marrée) :

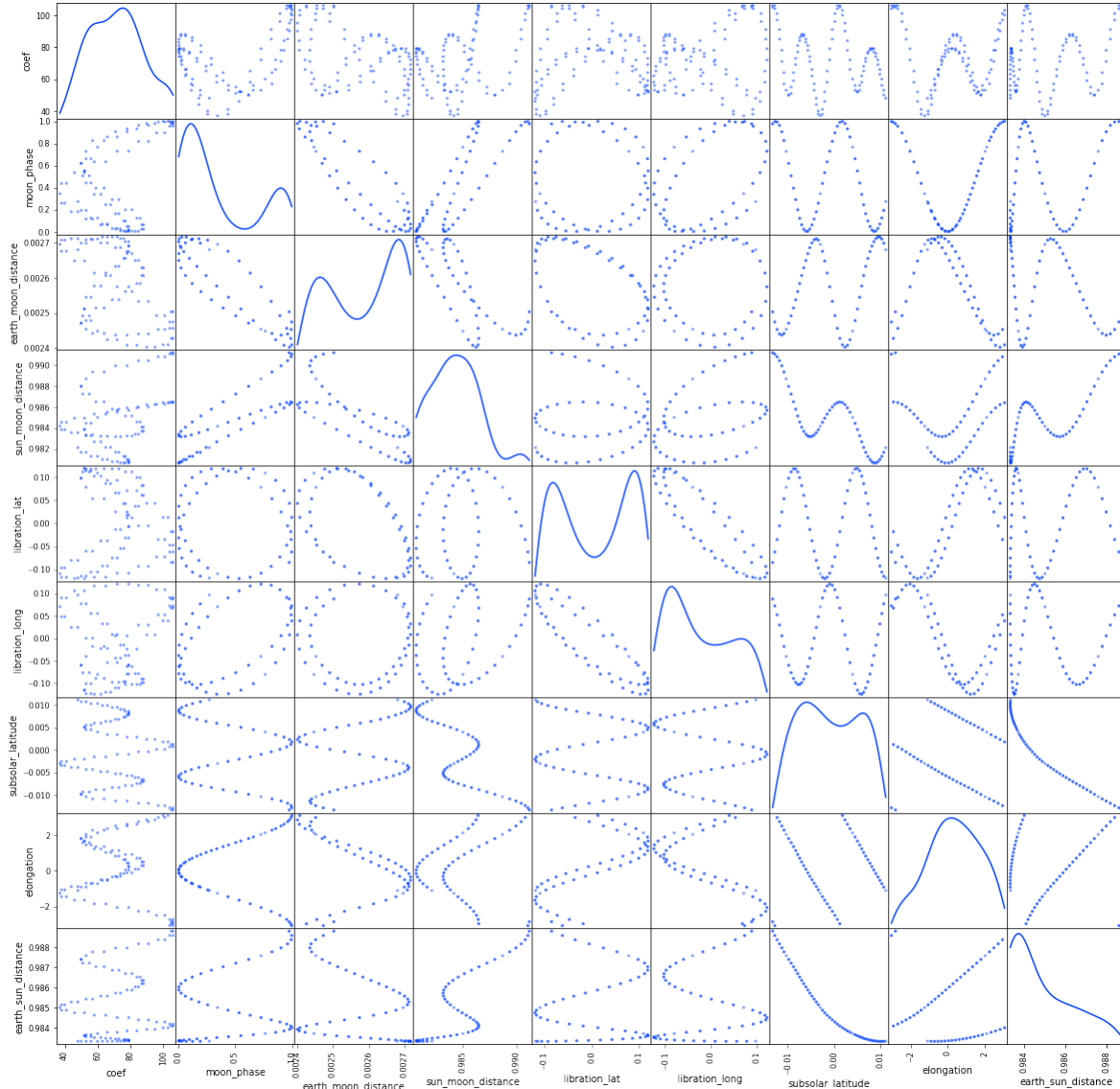


FIGURE 2 –

En regardant la première ligne de la matrice ci-dessus, on constate une forte corrélation entre les coefficients de marées et les différents variables ajoutées.

1.3 Etude d'importance : Quelle variable permet d'expliquer plus de variance dans l'espace de donnée

Toutes les variables semblent avoir une égale importance d'après le graphe ci-dessous :

On constate que presque toutes les composante ont une égale importance et donc doivent toutes être conservées.

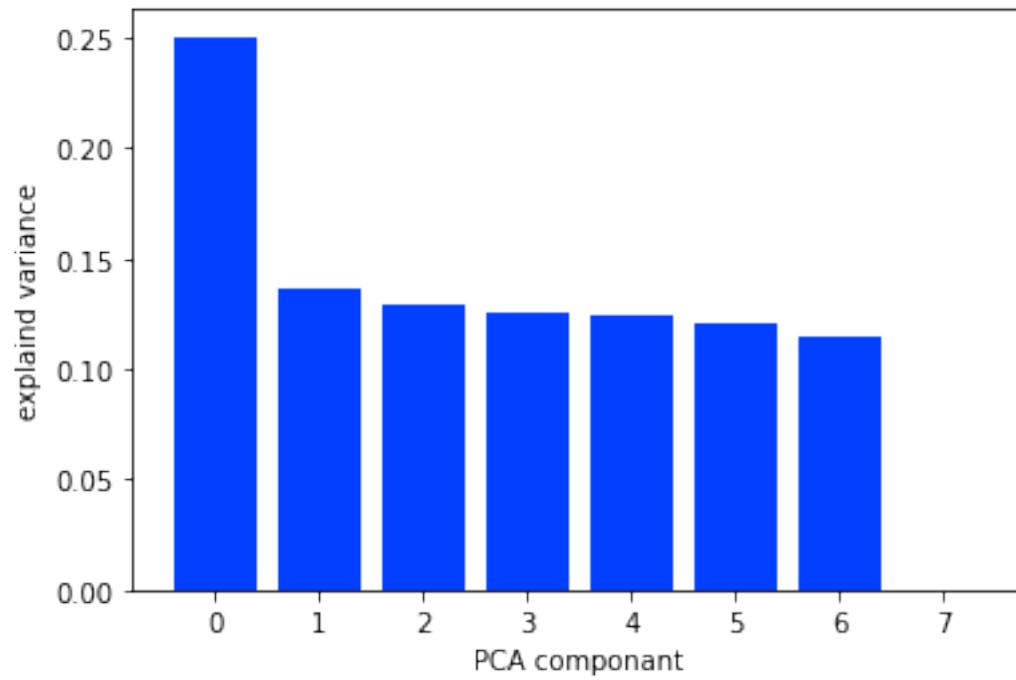


FIGURE 3 – La variance expliquée par chaque composante principale

1.4 Prédiction des coefficients en exploitant uniquement les grandeurs physiques

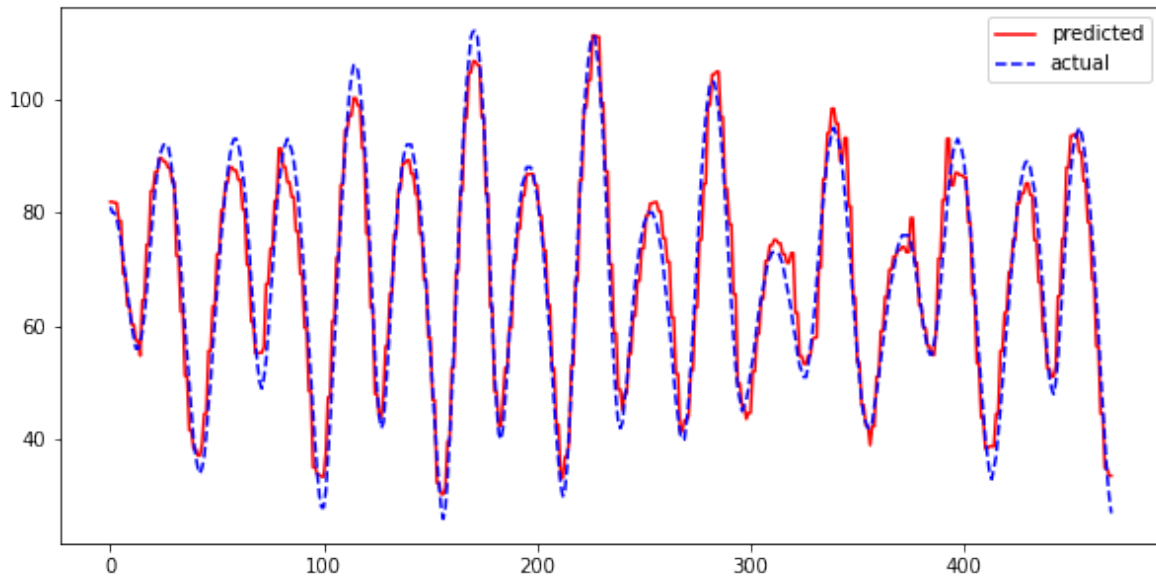


FIGURE 4 – Résultat de prédiction des coefficients de marée de l'année 2021 avec les grandeurs physiques.

Les résultats de prédiction sont meilleurs comparés aux performances atteintes en exploitant uniquement les variables temporelles. En effet, on obtient une erreur moyenne quadratique de :

$$RMSE(coef_predicted, coef_actual) = 4.92$$

En exploitant la réduction moyenne de l'impureté dans la construction des arbres du modèle des forêts aléatoires on a pu déterminer les variables les plus importantes (plus significatives que les autres en terme de prédiction du coefficient de marée). La figure ci-dessous permet d'identifier la variable la plus importante :

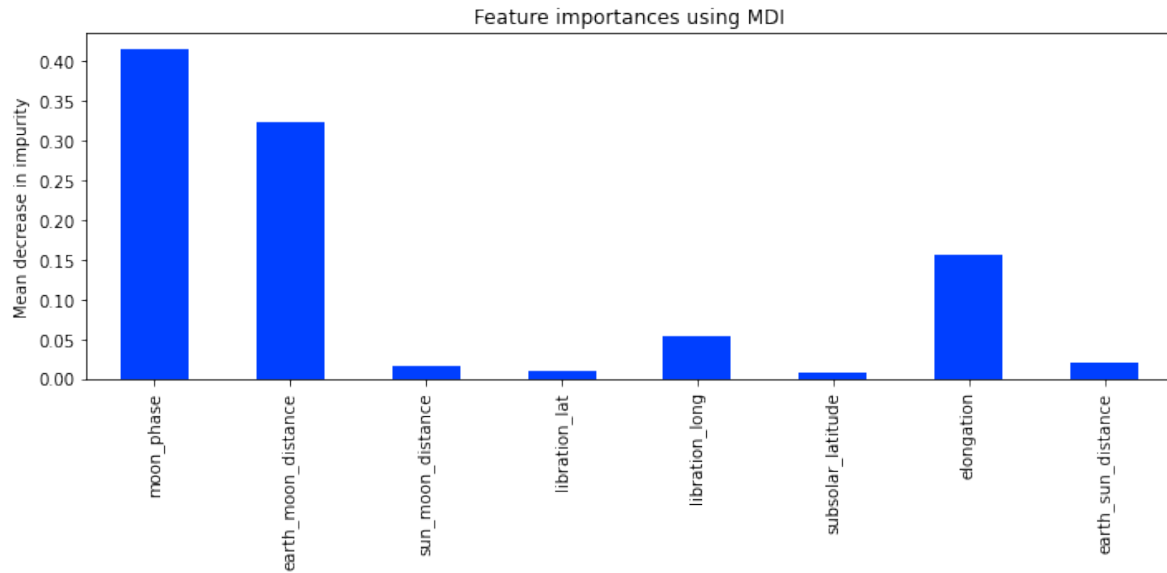


FIGURE 5 – L'importance des variables

On constate que la phase de la lune est la plus importante. Ce qui a été prévu.

2 L'interface WEB

Dans cette partie nous représenterons les différentes pages de notre interface WEB, que nous avons conçu jusqu'à maintenant. Il nous reste encore la partie Back-end du site afin que les différentes fonctionnalités fonctionnent correctement.

2.1 Home page

Cette page représente la page principale de l'interface WEB. Cette page permet au visiteur d'accéder aux différentes fonctionnalités du site WEB, principalement celle correspondantes à la prédiction des coefficients de marées.

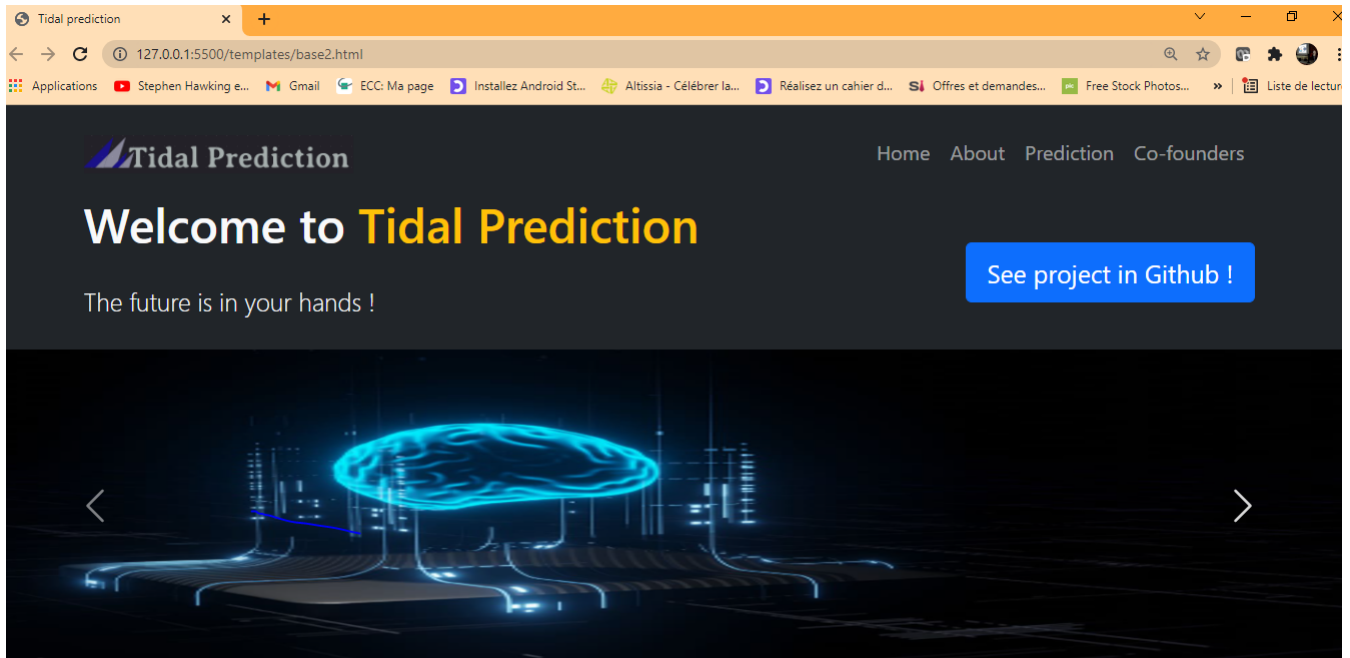


FIGURE 6 – Home page

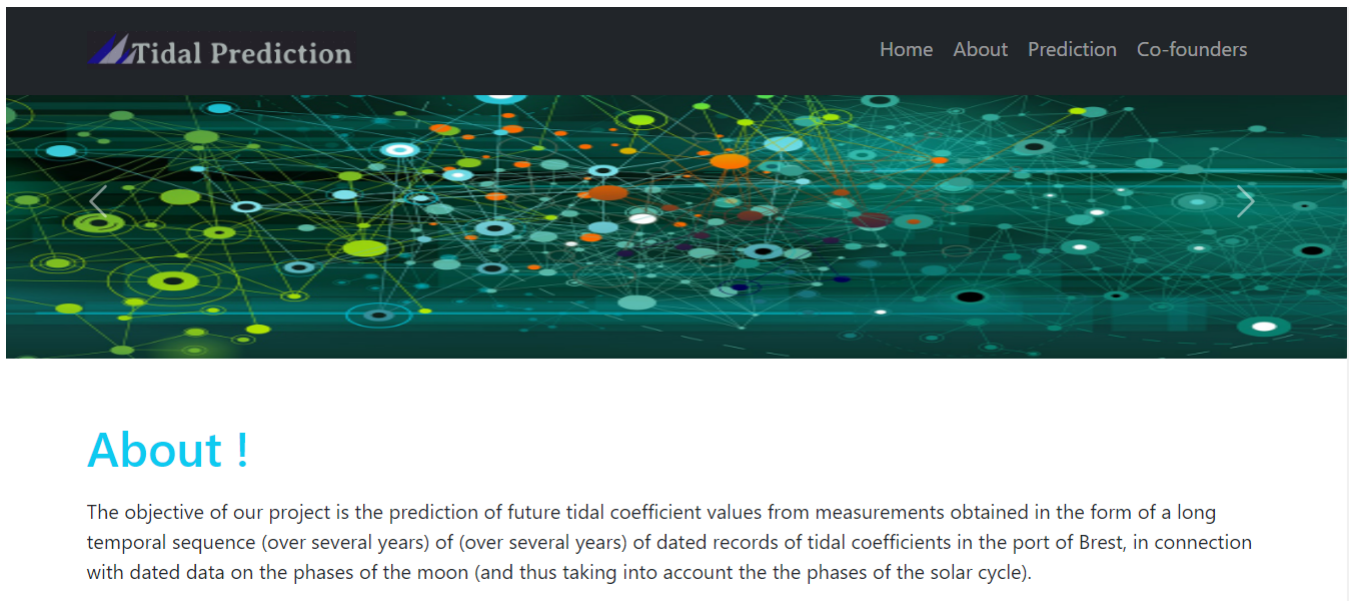


FIGURE 7 – Home page

2.2 Page de prédiction

Dans cette page, nous avons regroupé les fonctionnalités qui permettent à l'utilisateur de réaliser sa prédiction et visualiser les résultats. Cella sur un horizon qu'il va sélectionner selon une date de début et une date de fin.

Date	Min coefficient	Max coefficient
20/11/2021	60	67
21/11/2021	47	53
22/11/2021	41	43
23/11/2021	40	41

FIGURE 8 – Page de prediction

2.3 Co-founders page

Dans cette page nous représentons les membres de l'équipe qui a travaillé sur ce projet.

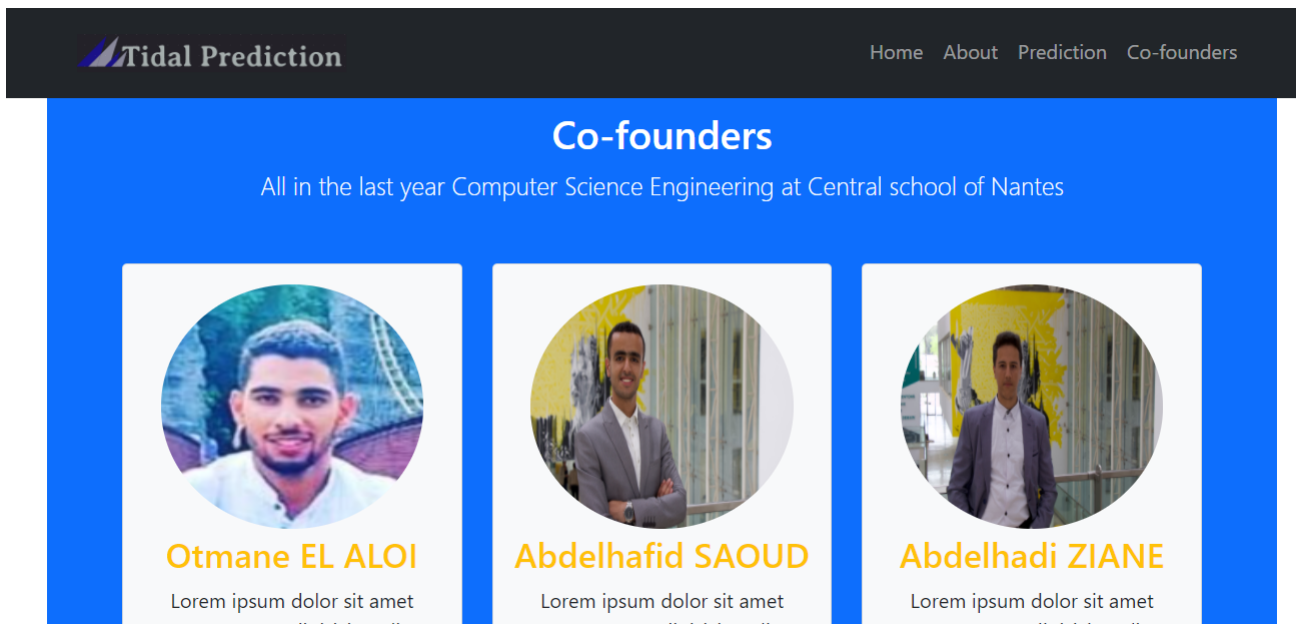


FIGURE 9 – Co-founders page

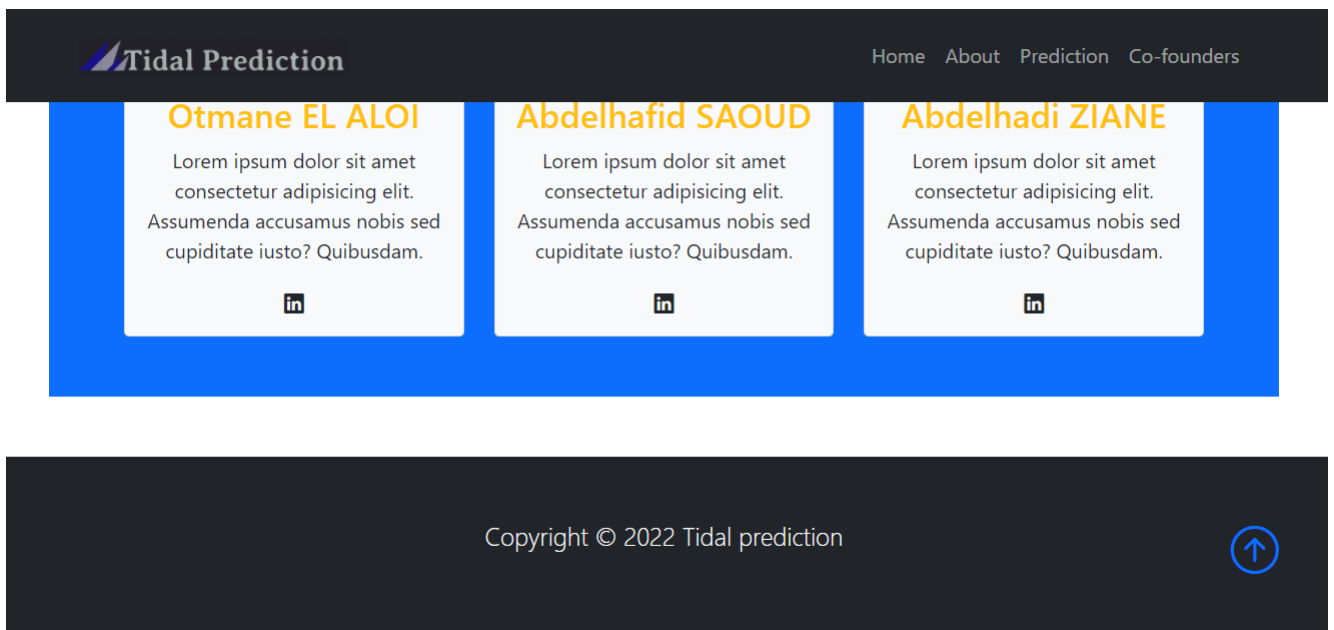


FIGURE 10 –

3 Planning

3.1 Planning de la semaine en cours

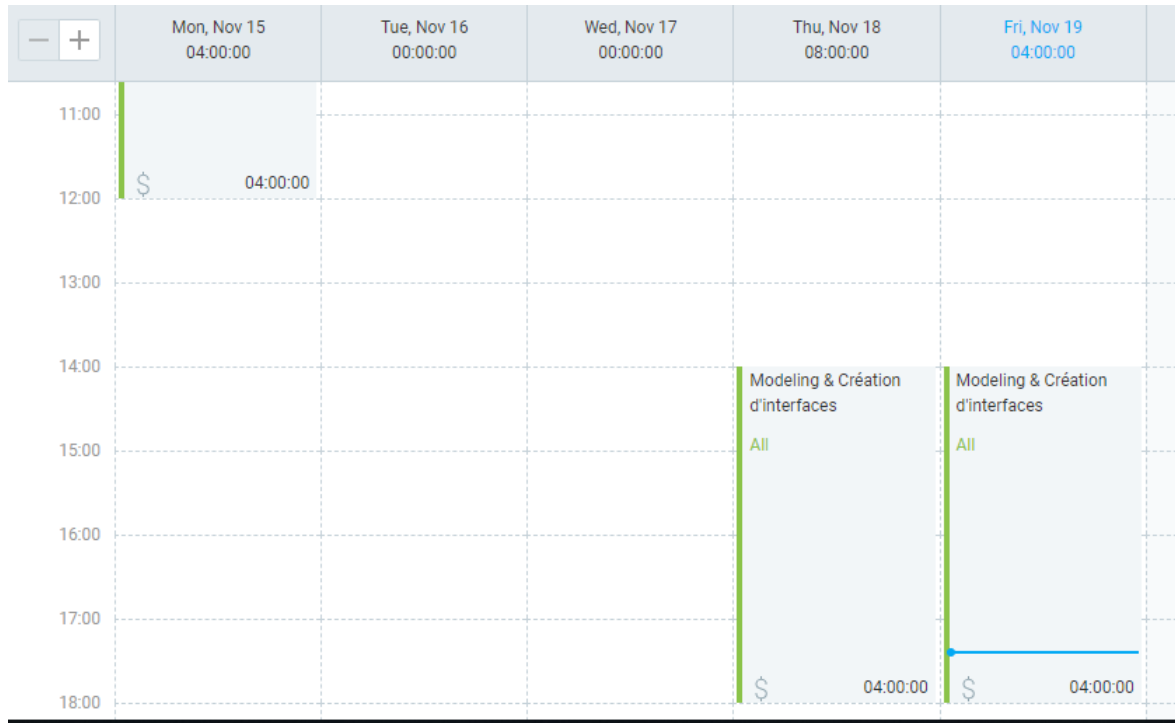


FIGURE 11 – Planning de la semaine en cours

3.2 Planning de la semaine prochaine



FIGURE 12 – Planning de la semaine prochaine