Résumé de chaque jour

**16/10/2023** **:** Premier jour à l’entreprise, rendez-vous avec Laurence FROISSARDEY, présentation de l’entreprise et de ses différents services ainsi que ses outils de travail, visite des locaux avec Fred.

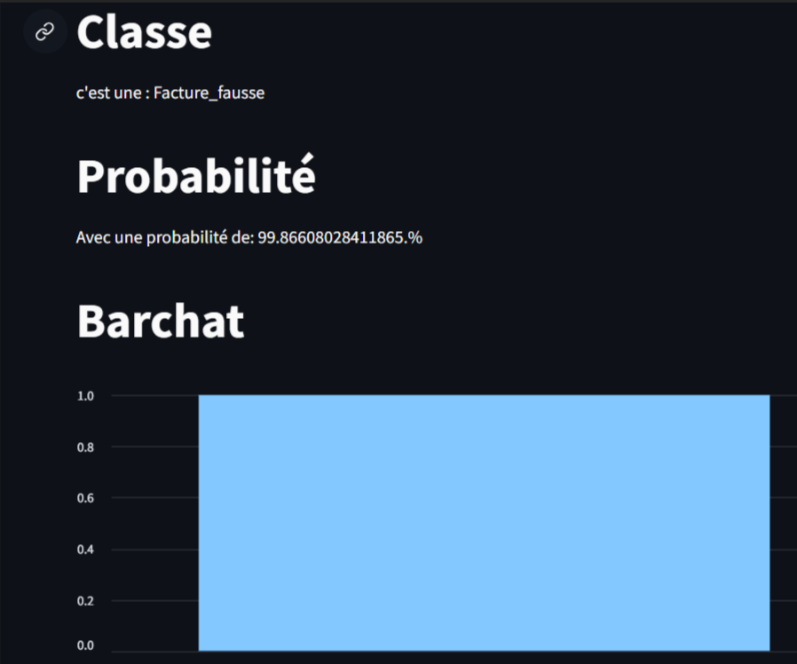
Présentation des différents projets et des scripts qui ont été mis en place avec Ben.

Mes première taches mon été attribué :

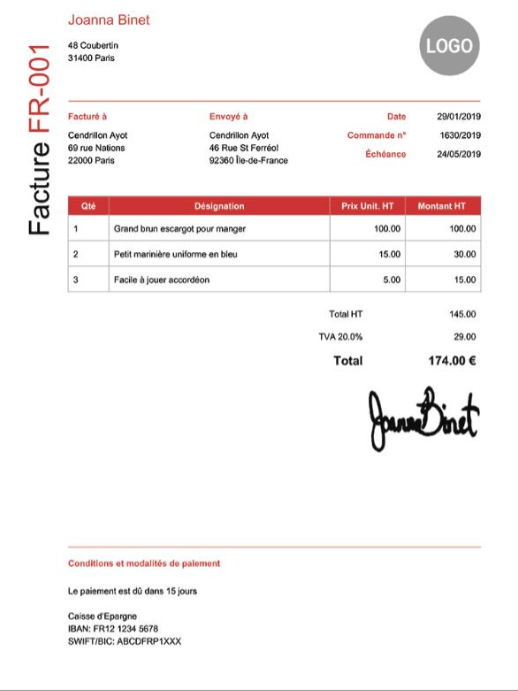
* Etudié ton environnement
* Volet de recherche des cas de fraude en assurance
* Comprendre les algorithmes déjà mit en place et en cours
* Implémentation des algorithmes pour lutter contre la fraude

**17/10/2023 :** je continue mes recherches sur la fraudes et test le modèle streamlit.

**J’ai donc fait des tests et voici les résultats :**



**Sur ce type de facture :**



**18/10/2023 :**

Recherches sur les différentes fraudes existante.

Réunion avec Ben à 10h30 dans là qu’elle on a pu voir les différentes recherches que j’ai faite quelques précisions ont été ajouter à cela.

J’ai par la suite effectué des recherches plus approfondies sur les réseaux de neurone notamment CNN et RNN ainsi que les différentes méthodes comme MAX POOLING ou encore STANDARDSCALER.

Cette partie est importantes avant de toucher aux scripts pour vraiment comprendre la théorie.

**19/10/2023 :**

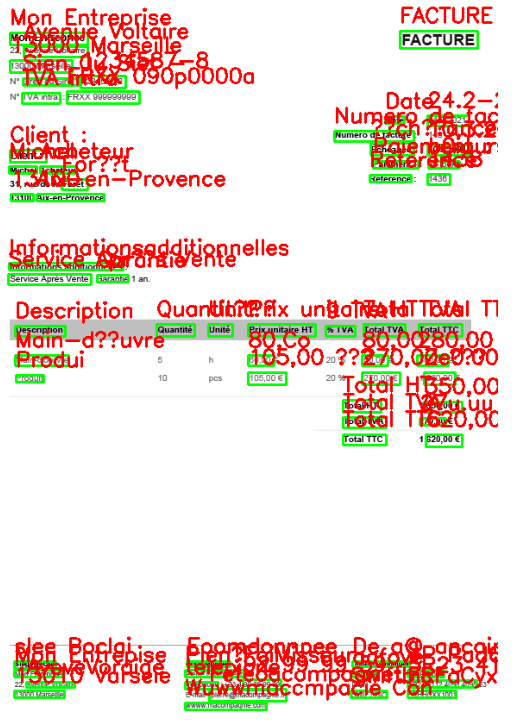
J’ai continué mes recherches sur les réseaux de neurone et je me suis penché sur les différents modèle OCR python qui pouvais exister telle que pytesseract, easyOCR et TrOCR. J’ai aussi installé vsCode et j’ai eu accès aux différents logiciels comme gitlab, guacamole et un GitHub partager avec ben.

Mais aussi au site Jira qui permet de faire des tickets lorsque l’on a n’importe qu’elle demande niveau informatique.

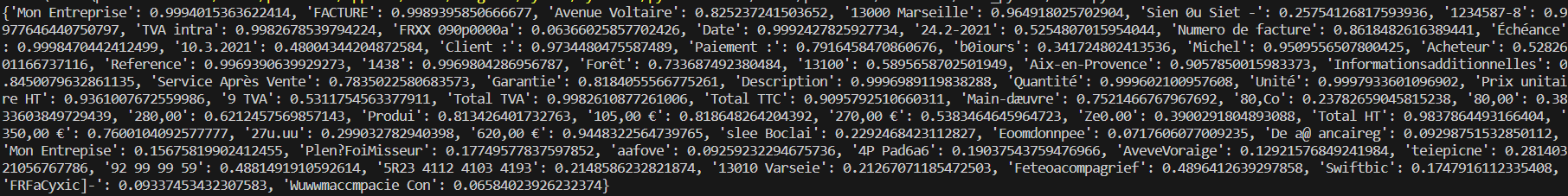
**20/10/2023 :**

Je développe un script qui permet d’utiliser l’OCR sur une facture et qui repère les différents mots et crée des bouding box autour de chaque mot ensuite grâce à la bibliothèque openCV on trace le rectangle qui a été délimité.

Voici un premier résultat :

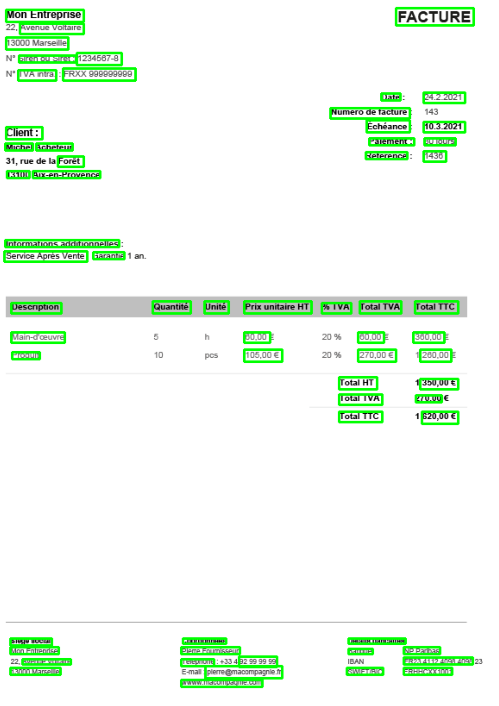


Et ressort en sortie :



Un dictionnaire avec en clé les différents éléments texte trouvé et en valeur la prob.

Sans les bounding box un résultat qui nous permet de voir se que le modèle repère :



**23/10/2023 :** recherches sur les différentes couches de neurones existantes (convolution, pooling, fully-connected(dense)…).

**24/10/2023 :** préparation d’une présentation d’un modèle d’intelligence artificielle et d’une API que j’ai mit en place à l’école

Lecture des différents scripts que Ben ma donné

Recherche de dataset de facture, pour remplir la BDD

**25/10/2023 :**

Présentation sur un projet que j’ai réalisé, du vocabulaire que j’ai trouve sur les réseaux de neurone.

Présentation de la maquette API mise en place par Ben.

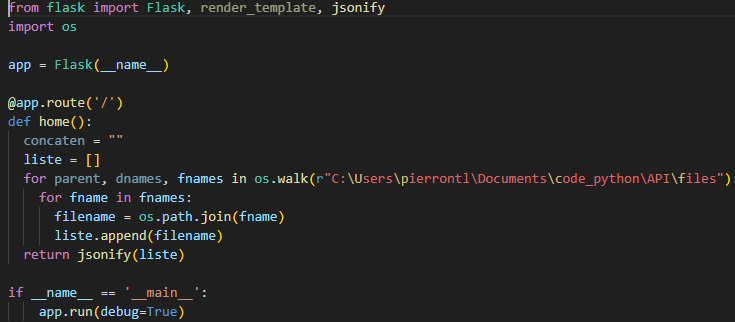
**26/10/203 :** présentation des différentes API, Flask et FastAPI.

Recherche sur une source de données pour alimenter le modèle.

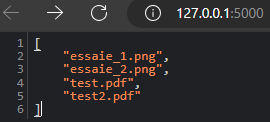
Préparation veille technologique sur les différentes avancé IA dans le domaine de la fraude.

Création d’une API qui permet de lire les différents fichiers enregistrer en local et qui permet d’afficher la présence de ses fichiers sur une page.

Code :



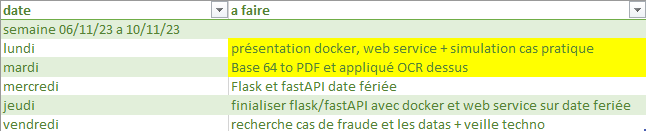
Résultat :



**27/10/2023 :**

Présentation veille techno et du code que j’ai mit en place pour l’API

**07/11/2023 :** établissement des taches de la semaine avec Ben voici le tableau :



Les éléments en jaune sont les taches que j’ai accomplies.

Le but pour base64 to PDF étais de crée une nouvelle fonction dans le fichier fonction ou étais recensé toutes les autres fonctions pour que le main.py fonctionne bien.

Le script est maintenant capable de prendre en entrée un PDF base64 de le convertir en PDF puis en image pour qu’il soit traitable par l’OCR.

Cette entrée est sous format JSON et ressort le résultat sous format JSON, True si trouve date fériée false sinon.

**08/11/2023 :**

Appliqué flask au code j’ai créé précédemment, le but étais d’avoir deux API une qui traite un PDF au format json donc elle prend en entrée un PDF et le transforme en base64 et une autre api qui permet de recevoir cette base64 et d’appliquer différents critères dessus.

**09/11/2023 :**

J’ai donc par la suite créé la deuxième API qui permettais d’appliquer les différents critères sur une facture.

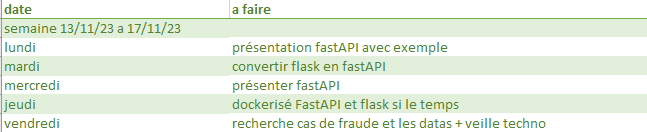
Présentation de docker et webservices.

**10/11/2023 :**

Je me suis initié a fastAPI crée un code tout simple pour comprendre la spécificité de l’API notamment avec le /docs qui permet de répertorier tous les Endpoint que l’on a créé.

J’ai aussi fait une veille techno sur chatgpt et DALL-E ainsi que la recherche de data.

J’ai établi le planning de mes taches pour la prochaine semaine qui est le suivant.



**13/11/2023 :** j’ai avancé sur la présentation de fastAPI que j’ai terminée et j’ai commencé à convertir l’api que j’ai construit Flask en fastAPI.

**14/11/2023 :** j’ai continué de développer l’API fastAPI j’ai tout mit dans un seul fichier je n’ai pas encore conçu les deux API bien distinct.

Je me forme à docker.

**Pour mercredi 15/11/2023 a vendredi 17/11/23 :** j’ai fait un fichier readme.md qui permet de comprendre ceux que j’ai pus mettre en place il explique l’utilisation de l’api fastAPI et de son évolution au cours du temps.

**Du lundi 25/11/2023 au vendredi 01/12/2023 :** j’ai continué sur le développement distinct des deux API que j’ai finie et que j’ai mit sur docker on essaie maintenant de le faire marcher chez Ben mais quelques problèmes ont eu lieu ce qui freine un peut le processus.

J’ai trouvé un contact sur LinkedIn très intéressant qui nous permet d’avancer sur les différentes méthodes de fraude employé par les fraudeurs, un deuxième contact est donc ressorti de cette réunion qui est l’entreprise de CTMS avec qui j’ai planifié un rendez-vous à 15H.