Analyse des plateformes

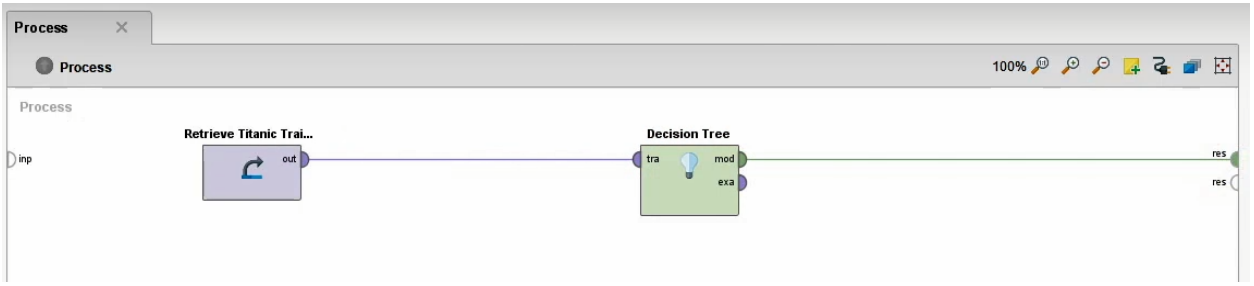
Orange3

Se reporter au Cahier des Charges

RapidMiner

Utilisé en data science, surtout pour la préparation de données.Rapidminer permet de faire des liens entre des valeurs sur des tableaux de base de données. Peut faire des prédictions pour compléter ces tableaux, trouver des corrélations.

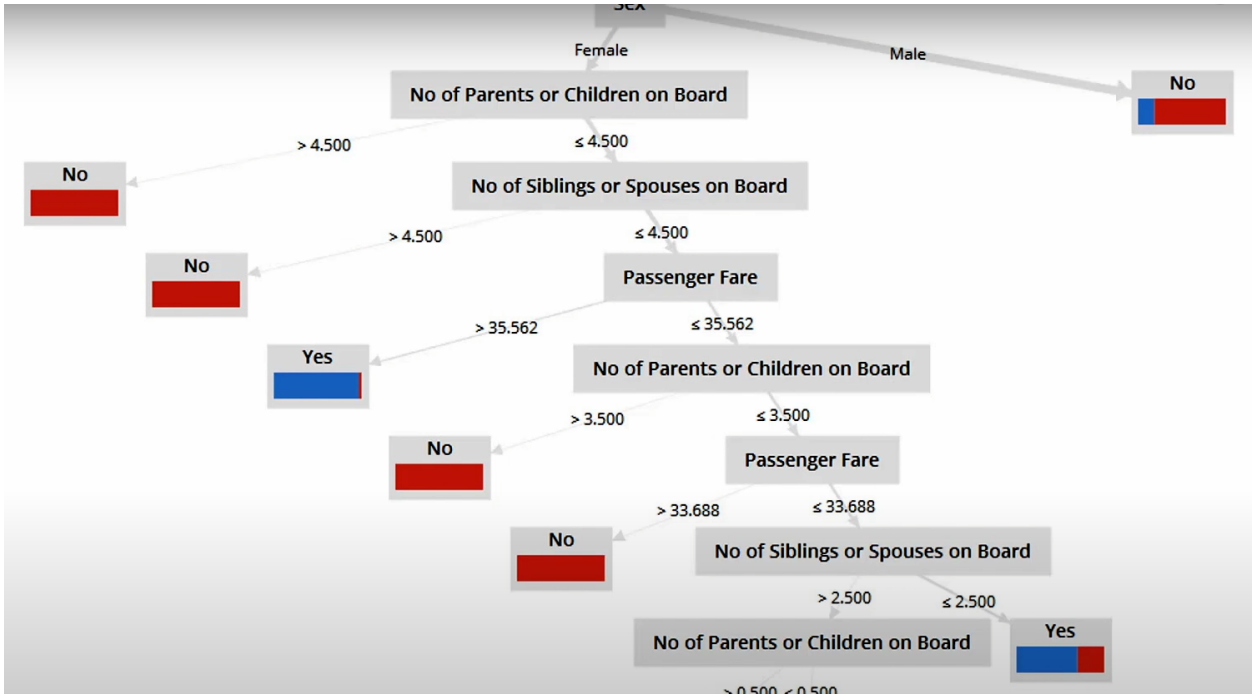
->Analyse statistique, data mining et prédictions.
Fonctionne sur le principe d'arbres de décision, pratique d'utilisation avec un codage par visualisation graphique. Permet de faire des choix basés sur une suite d'étapes précises à connecter entre elles.



Il y a des systèmes de validation croisée, se basant sur des modèles d'IA pour réaliser des estimations de fiabilité sur les données.

Les outils de RapidMiner permettent de générer un modèle prédictif à partir de ces étapes :

- La préparation des données ;
- La sélection des variables ;
- La définition des éléments de prédiction ;
- La création du modèle.



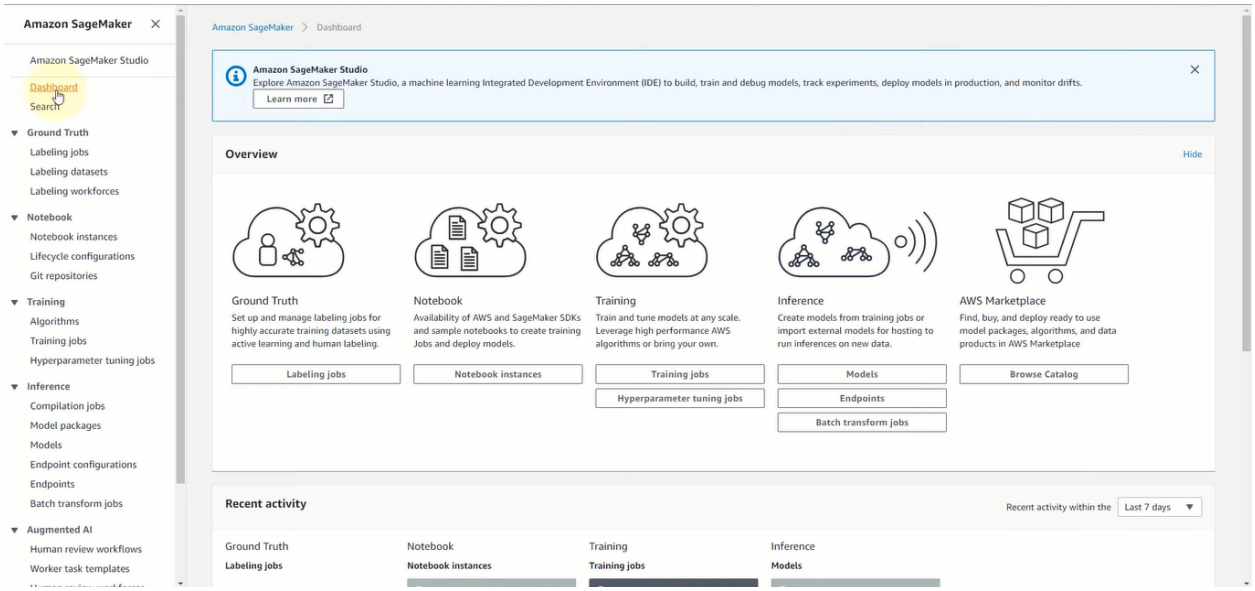
Arbre de décision permettant de faire le lien entre certaines données et le résultat.

Ce logiciel est parfait pour l'analyse de données, il permet surtout d'étudier des corrélations entre certains types de données. Il ne permet cependant pas la prédiction type LSTM qui nous intéresse, bien que ce soit un bon outil pour en faire la validation.

Amazon SageMaker

Amazon SageMaker couvre l'ensemble de ce que l'on peut faire en Machine Learning et en Apprentissage Automatique. Allant de l'étiquetage des données à la création d'instances en passant par la possibilité d'exploiter des notebook déjà existant, SageMaker permet de réaliser du Machine Learning à toutes les échelles et de régler des paramètres en détail, mais aussi d'acheter ou vendre des modèles.

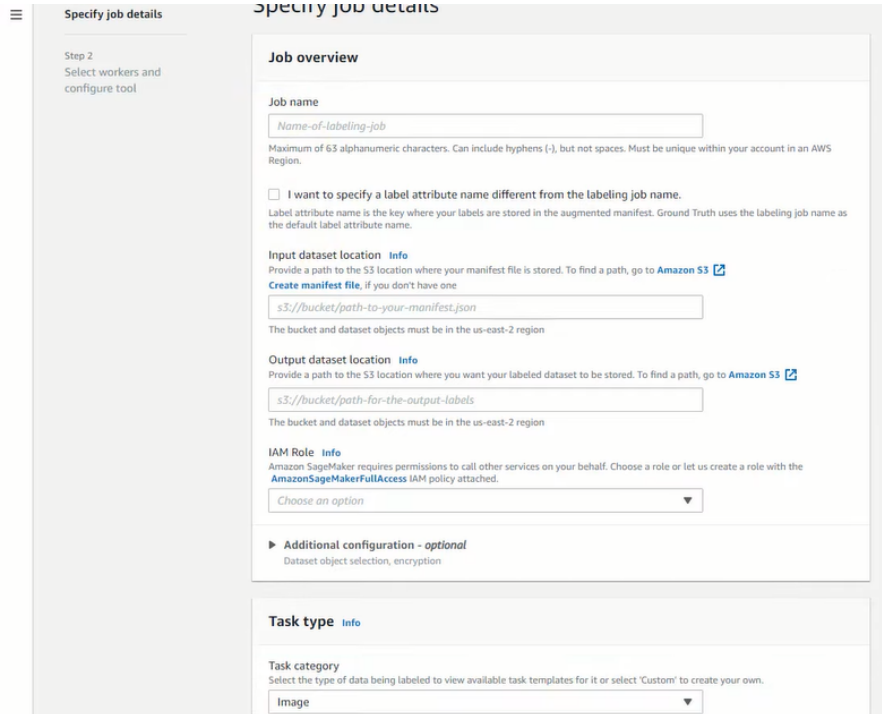
La page d'accueil présentant ces différentes options est affichée ci-dessous.



Prenons l'exemple de l'étiquetage de données, qui est représentatif de la manière de fonctionner de SageMaker. L'interface est principalement constituée de paramètres à définir en cochant ou en remplissant des cases. Les choix se font ainsi en quelques clics, ce qui fait que la prise en main de cette plateforme est assez rapide.

Cependant, le vocabulaire utilisé est assez avancé. Ceci couplé au manque de dessins ou schémas explicatif en fait une plateforme utile pour des personnes ayant déjà des notions en IA. Pour illustrer ceci, voici ci-après un capture d'écran de la partie étiquetage de données.

SageMaker semble donc assez peu adapté pour les débutants.



Néanmoins, cette limite signifie aussi que les types de tâches que peut réaliser cette plateforme sont très diverses : classification pour une ou plusieurs étiquettes, clustering, segmentation sémantique (segmentation au pixel près) enfin vérification des classes …

Le tout en quelques clics.

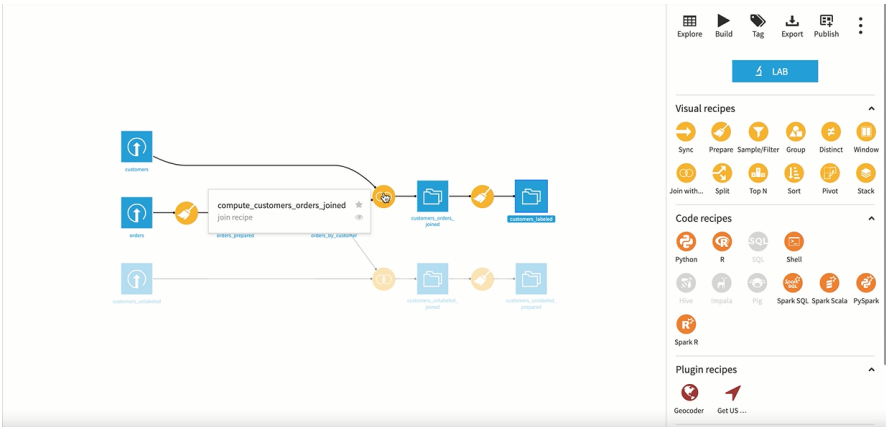
SageMaker fonctionne en créant des Jupyter Notebook afin que l'on puisse coder, télécharger nos données, etc. Il y a aussi à disposition des algorithmes usuels réalisant différentes tâches (clustering kmeans, etc)

De part la possibilité de partage des modèles, SageMaker semble parfaitement adapté pour des fins professionnelles, mais son utilité dans le cadre de l'apprentissage est discutable.

Dataiku

Dataiku se distingue des autres plateformes de par son esprit collaboratif au sein d'un même environnement, permettant de réunir les Data Scientists, analystes et opérateurs.

Pour les analystes, elle se présente comme une interface visuelle interactive au sein de laquelle il est possible de pointer, cliquer, et développer en utilisant des langages comme SQL (voir image ci-dessous). Ainsi, il est possible de confronter des données, modéliser, relancer les workflows, visualiser les résultats, et obtenir des insights sur demande. Ces fonctionnalités permettent aux Data Analysts d'augmenter leur efficacité.



Pour les Data Scientists, la plateforme permet de préparer et de modéliser les données en quelques secondes. A noter également que l'interface est entièrement personnalisable, ce qui permet d'automatiser les tâches.

Pour les opérateurs, la plateforme retire l'inquiétude liée à l'utilisation de plateformes multi-technologiques. Elle permet de coordonner le développement et les opérations grâce à l'automatisation du workflow, la création de services web prédictifs, et la surveillance du statut des données et des modèles au quotidien.

En termes d'utilité, la plateforme touche à peu près tous les domaines liés à la Data Science. L'interface est constituée de schémas permettant une bonne compréhension, tout en restant assez technique au niveau de la prise en main, mais permettant in fine des raccourcis et des généralisations non négligeables.

Alteryx

Alteryx est un progiciel dédié à l'automatisation des tâches. Ce logiciel propose énormément de fonctionnalités différentes, notamment en machine learning, cependant il est surtout axé sur l'analyse prédictive et est donc principalement utilisé en data science. Alteryx étant un produit utilisé par de nombreuses entreprises dans le monde, il se veut très facile d'accès et de prise en main. Un débutant en informatique doit pouvoir réaliser des tâches complexes en machine learning et data science. De ce fait de très nombreux presets permettant de satisfaire un maximum de tâches courantes sont disponibles ce qui rend sa prise en main très facile, notamment grâce à l'ergonomie et l'esthétisme du logiciel. Ce dernier est également utilisable par des personnes expérimentées en informatique, mais ces dernières se trouveront peut-être quelque peu ralenties par le design logiciel volontairement accessible.

Au niveau purement technique Alteryx est capable d'effectuer l'ensemble des tâches les plus classiques de data science, qu'il s'agisse d'analyse prédictive ou de gestion de base de données, en passant par des possibilités plus uniques comme l'analyse géospatiale.

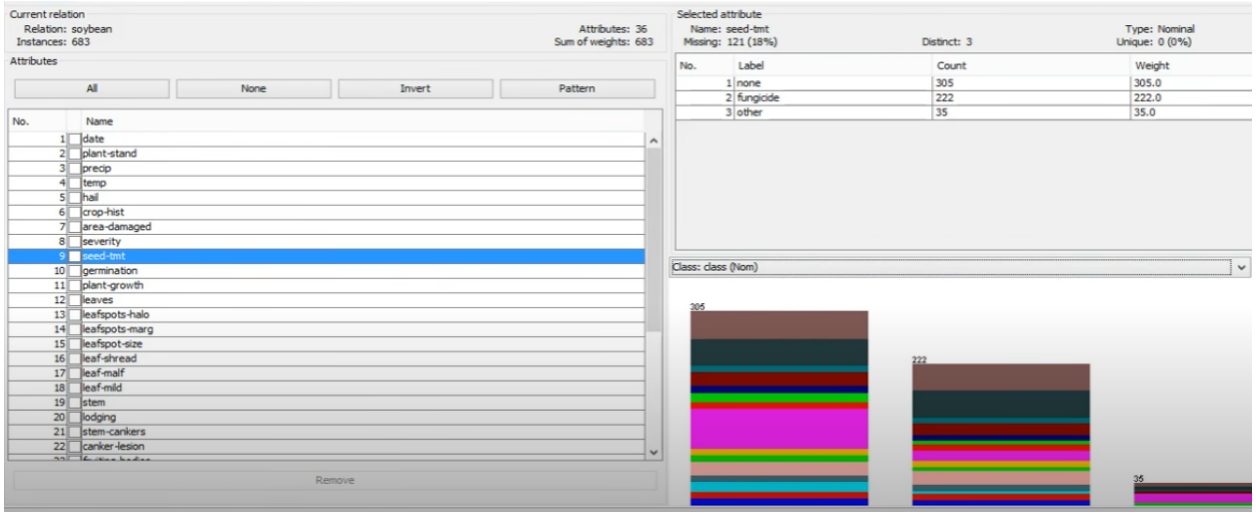
En résumé, cet outil n'est pas adapté à proprement parler à la création de RNN. Cependant sa grande accessibilité et sa communauté dynamique en font un parfait outil pour découvrir le machine learning et les data science.

KNIME

KNIME est un logiciel libre et open-source d'analyse de données utilisant une interface graphique similaire à LabView. Ce logiciel permet l'intégration de divers langages de programmation et outils ainsi que la création automatique de comptes-rendus. KNIME comprend un ensemble d'outils pour l'apprentissage automatique et l'exploration de données par le biais d'une interface de workflow modulaire. La plateforme se veut facile d’accès mais permet tout de même une forte personnalisation par le code. Ce logiciel, bien que axé sur les data science permet la création de réseaux de neurones LSTM et propose même un cours sur la reconnaissance d’émotions basé sur ces derniers. Son interface graphique permet de travailler rapidement sur la création de réseaux de neurones et en fait un bon outil pour les débutants. Sa communauté est cependant moins développé que celle du progiciel Alteryx

WEKA

Collection d'algorithmes pour l'IA dans un logiciel JAVA Open Source.
Outils qui propose du data mining, machine learning, preprocessing, classification, regression, clustering, visualisation etc..



Travaille principalement sur les tableaux de données type SQL. Pas de multi-relationnel. Donc sans doute moins intéressant pour l'utilisation d'un réseau de neurones LSTM.
Fonctionne avec:
-Arbres décisionnels
-classifieurs basés sur des instances
-Support machines vectorielles
Schéma implémentés:
K-means, EM, cobweb, X-means, FarthestFirst.
Propose des interfaces graphiques pour une meilleure compréhension, avec clustering.