

Proposition de Projet Pluridisciplinaire pour la formation MAGE Paris Descartes

Information concernant l'encadrant

[Prénom Nom] Etablissement et poste : Soror Sahri, Université Paris Descartes

Email : soror.sahri@parisdescartes.fr

Téléphone :

Description générale du projet

Intitulé du projet : outil d'évaluation de la qualité de données dans un système NoSQL

Contexte :

Les données augmentent en volume et en vitesse d'arrivée. Les traitements (gestion et fouille de données) deviennent complexes. Alors que ces Vs (volume et vitesse) sont relativement bien définis et peuvent être facilement mesurés, la véracité n'est pas bien définie et difficile à mesurer.

En effet, si les données sont inexactes, imprécises, incomplètes ou incohérentes, tous les résultats analytiques sont sans signification et peu fiables. Ainsi, la validation de la qualité des données, tout particulièrement leur véracité, est très importante avant et en temps d'interrogation de données.

Ce projet consiste à implémenter une méthode existante sur l'évaluation de la véracité des données en tenant en compte des différentes dimensions de la qualité de données. Il existe plusieurs dimensions de la qualité de données : la consistance et cohérence des données, leur complétude, leur actualisation (ang. freshness) des données), etc.

Réalisations attendues :

- une application permettant d'évaluer la qualité des données selon une méthode existante.

L'application aura en entrée un jeu données, ainsi que le modèle des métriques de qualité ; en sortie : les mesures de la qualité sont évaluées selon chaque dimension de la qualité. L'évaluation est effectuée sur des données volumineuses et distribuées stockées dans un système distribué à grande échelle, notamment Spark.

Niveau requis* : M1 / M2

** si M2, précisez s'il s'agit de la poursuite par une même équipe d'un sujet déjà abordé l'année précédente et indiquez l'équipe*

Description détaillée du projet

Description des tâches fonctionnelles :

- Etude de la méthode d'évaluation de la qualité des données. Cette étude consiste à comprendre les différentes dimensions de la qualité des données (cohérence, complétude, etc), étudier le modèle de l'évaluation de la qualité et l'améliorer éventuellement.
- Rédaction de la documentation fonctionnelle et technique

Description des tâches techniques :

- prise en main du système Spark
- implémentation de la méthode d'évaluation des données.

Proposition de Projet Pluridisciplinaire pour la formation MAGE Paris Descartes

Information concernant l'encadrant

Laurent Grizon – IFPEN – Chef de Projet - et Mathieu Piccin xxxx

Email : laurent.grizon@ifpen.fr mathieu.piccin@gmail.com

Téléphone : 06 10 19 51 92

Description générale du projet

Intitulé du projet : développement de l'application InfoMap et gestion du projet

Contexte :

La société SocialWork souhaite faire développer l'application InfoMap destinée à géoréférencer des données et les cartographier de façon interactive pour disposer de tableau de bord dans le domaine social.

Objectifs :

- L'objectif de l'application de cartographie « InfoMap » est de créer interactivement via une carte 2D un jeu de données comprenant des points auxquels sont liées d'autres données de nature variable (texte, nombres, %, etc ...), de visualiser l'ensemble de ces informations sur cette carte 2D et de sauvegarder l'ensemble des données. Toutes les fonctionnalités décrites dans ce document seront accessibles via la carte.

- Il s'agit aussi de produire les documents associés à la gestion du projet (chiffrage-planification)

Réalisations attendues (e.g. site web, application mobile) :

- application informatique
- gestion du projet dont chiffrage, planification

Niveau requis* : M1 ou M2

** si M2, précisez s'il s'agit de la poursuite par une même équipe d'un sujet déjà abordé l'année précédente et indiquez l'équipe*

Description détaillée du projet

Description des tâches fonctionnelles :

- Planification minutieuse du projet
- Rédaction de l'ensemble des documents de pilotage
- Rédaction de l'expression des besoins, des spécifications et analyse

Description des tâches techniques :

- Conception et développement de l'application
- Réalisation des tests et de la recette
- Validation de la faisabilité (performances et ergonomie)

Compétences fonctionnelles et techniques requises (e.g. langage de programmation, outils de conception, bibliothèques) :

- Python
- NetCDF

Proposition de Projet Pluridisciplinaire pour la formation MAGE Paris Descartes

Information concernant l'encadrant

[Prénom Nom] Etablissement et poste : François Role, Paris Descartes, Maître de Conférences

Email : francois.role@parisdescartes.fr

Téléphone : /

Description générale du projet

Intitulé du projet : Objets connectés et reconnaissance d'images pour assister les handicapés visuels

Contexte et objectif : le projet part du constat que, dans la vie de tous les jours, les technologies avancées profitent encore très peu aux handicapés notamment aux handicapés visuels qui utilisent aujourd'hui à peu près les mêmes techniques d'assistance que celles qui ont été élaborées il y a un siècle (chien-guide, canne). Ceci est notamment dû au fait que cette population ne représente pas un marché économique considéré comme intéressant. Des techniques peu coûteuses et très puissantes sont cependant disponibles aujourd'hui.

L'objectif du projet est donc de développer des outils d'assistance visuelle (détection objets, de visages) pour des personnes aveugles ou déficientes visuelles. Tous les développements seront faits en *Open Source*. Ces outils développés seront mis gratuitement à la disposition de la communauté concernée. Le projet demandera du travail mais sera motivant car il a pour objectif d'aboutir à des résultats concrets grâce notamment à une suite en master2 (voir ci-dessous).

Travail à réaliser : une caméra portative de type webcam sera connectée à un ordinateur ultra-léger de type Raspberry PI. Les images récupérées sur le Raspberry seront envoyées pour traitement à un serveur distant exécutant un modèle de reconnaissance d'images. Le résultat de l'analyse sera retransmis à l'utilisateur sous forme textuelle non audio dans un premier temps.

Le projet est prévu pour avoir une suite en deuxième année, où plusieurs scénarios alternatifs seront explorés : utilisation d'une caméra "intelligente" permettant d'effectuer des traitements temps réel en local ; utilisation de machines situées sur un Cloud. Des contacts sont déjà pris et seront poursuivis en cours d'année avec l'équipe AWS France pour voir si la deuxième partie du projet pourrait utiliser des technologies de type *DeepLens*.

Niveau requis* : M1 / M2

* si M2, précisez s'il s'agit de la poursuite par une même équipe d'un sujet déjà abordé l'année précédente et indiquez l'équipe

Description détaillée du projet

Organisation

Le projet associera une équipe du master Miage et une équipe du master MLSD. L'équipe Miage se concentrera sur les aspects Web et déploiement tandis que l'équipe MLSD travaillera sur le développement des modèles d'apprentissage. Un soin particulier devra être apporté à la bonne définition et documentation des entrées sorties pour que les deux équipes puissent bien collaborer. Une vraie gestion de projet devra être mise en place et la coordination s'effectuera en utilisant des outils coopératifs avec l'aide de l'encadrant.

Les développements seront guidés par au moins une et peut-être deux personnes utilisatrices potentielles qui pourront suggérer des idées et apporter un retour.

Si des résultats présentables sont obtenus, des actions de communication pourront être menées.

Technologies à mettre en oeuvre et compétences à développer

- Programmation Python sur Raspberry PI.
- Compétences client-serveur pour développer un serveur Flask chargé de gérer les images venant de la caméra (gérer notamment du streaming video).
- Compétences en développement de modèles d'apprentissage.

Le projet permettra aux étudiants du master Miage de découvrir le *Machine Learning* et aux étudiants du master MLSD de s'initier aux objets connectés. Les deux groupes seront également sensibilisés à la question du handicap.

Proposition de Projet Pluridisciplinaire pour la formation MIAGE Paris Descartes

Information concernant l'encadrant

[Prénom Nom] Etablissement et poste : Gilles POIROT,
Architecte d'Entreprise et Logiciel,
THALES Services,

Email : gilles.poirot.parisv@gmail, gilles.poirot@thalesgroup.com

Téléphone : 0173321766 (Office) 0689712487 (Mobile)

Description générale du projet

Intitulé du projet : Zombies Apocalypse Simulation

Contexte :

Avec l'évolution des architectures de SI modernes vers une distribution massive (micro services, server less) il s'est avéré impossible de tester les applications de façon exhaustive à priori hors du contexte de production. Ces tests en production sont à la base d'une nouvelle discipline, le Chaos Engineering dans laquelle on utilise, entre autre, l'injection de "failure" et l'on regarde comment le système réagit afin de prévenir les problèmes avant qu'ils n'arrivent, ce qui sera inéluctable. Si au départ cette approche concernait plutôt les incidents techniques, elle s'étend naturellement aux problèmes humains pouvant affecter le bon fonctionnement de l'organisation. On parle alors de Zombie Simulation où des membres de l'organisation passe dans un mode "Zombie". Le Chaos Engineering a un lien fort avec le Disaster Recovery et l'on peut comparer Zombie Apocalypse à du tes de Disaster Recovery d'un point de vue Business

Objectifs :

- Développer une application permettant de gérer des expérimentation de type Zombie Apocalypse au niveau de l'entreprise aussi bien au niveau de la planification et du suivit des ces expérimentations, que du contrôle lors de leur déroulement.

Réalisations attendues (e.g. site web, application mobile) :

- Application Mobile de contrôle,
- Application Web (ou mobile sur tablet) de planification

Niveau requis* : M1 / M2

En fait il y a potentiellement deux applications à réaliser, une plutôt "backend" et une "edge/mobile". L'application Mobile est peut être plus inhabituelle. Dans tous les cas il pourrait y avoir deux équipes et pourquoi pas une M1 et une M2

Description détaillée du projet

Description des tâches fonctionnelles :

- Etablir un Domain Model (Ubiquitous Language) selon l'approche DDD (Domain Driven Design),
- Définir des Objectifs et les motivations, stratégies pour les atteindre,
- Définir des User Stories exécutable (RTS / MVP)

Description des tâches techniques :

- Développement d'une application de gestion des simulations de Zombies (notamment leur planification: qui devient zombie, quand devient t-il Zombie, quel est son degrés de Zombie Attitude, quand redevient il en mode normal
- développement d'une application mobile pour les Zombies (pour pouvoir les assister dans leur comportement Zombie mais aussi enregistrer les actions qu'ils font ou ne font pas (appel rejetés, réunion absent, travaux non remis...))

Compétences fonctionnelles et techniques requises (e.g. langage de programmation, outils de conception, bibliothèques) :

- Langages de programmation: Swift pour iOS (et Kotlin si développement Android), Go côté serveur,
- Base de données Realm et aussi FireBase,
- Framework Firebase, Apache OpenWhisk,
- iOS, Xcode

Interface web interactive pour l'analyse d'audience de sites web

RGoogleAnalytics, Shiny

Contexte du sujet

R, le langage de programmation populaire pour l'informatique statistique, est un outil puissant pour analyser et tirer des conclusions à partir de données. Lorsque vous combinez **R** avec vos données **Google Analytics**, vous pouvez effectuer une analyse statistique et générer des visualisations de données pour mieux comprendre et améliorer votre entreprise.

Google Analytics est un service de mesure d'audience des sites Web. On parle de Web Analytics. L'objectif est d'apprécier la fréquentation. Mais au-delà du simple comptage, et c'est ce qui fait réellement sa force, l'outil propose de nombreuses possibilités d'études qui permettent de mieux comprendre le comportement des visiteurs. Ainsi, en identifiant les pages les plus regardées, les pages d'entrées, en analysant les trajectoires, en ventilant les accès selon les pays, etc., il nous permet de mieux organiser notre site de manière à répondre plus précisément aux attentes des internautes.

RGoogleAnalytics est un package R permettant d'accéder aux données Google Analytics via une API (interface de programmation applicative). La librairie nous donne accès aux informations stockées par le service, tant en termes de fréquentation brute qu'en termes d'analyses préprogrammées. Concrètement, nous spécifions les axes d'étude (les « dimensions » ou les « contextes » si l'on se réfère à la terminologie Business Intelligence) et les mesures (les « faits »), l'API fournit les tableaux correspondants. On pourrait par exemple extraire facilement le nombre de sessions (mesure) par pays (dimension).

Cette utilisation de type reporting est déjà intéressante en soi. Nous recourrons à R pour produire une interface web interactive en utilisant le package **shiny**, et des tableaux et graphiques avenants à partir des données fournies par Google Analytics. L'intérêt du dispositif est décuplé lorsqu'on se rend compte qu'en rapatriant les données sous R, un outil statistique puissant, il est possible d'initier des études approfondies de data mining.

Travail à faire

L'objectif du PPD est de développer une application Web client-serveur en utilisant le package Shiny développée par Rstudio et permettant d'exécuter une tâche d'analyse de données et de data mining pour analyser des données provenant de Google Analytics, et de visualiser l'information extraite depuis un navigateur. L'outil à utiliser pour construire l'interface sera le package Shiny, les packages et outils de data mining à interfacer seront les packages 'Rgoogleanalytics', 'wordcloud', ainsi que d'autres packages de data mining développés sous R. Le projet comporte cinq aspects principaux:

1. Extraction des données à partir de GoogleAnalytics
2. Prétraitement et transformation de données
3. Analyse exploratoire (statistiques et représentation graphique)

4. Affichage interactive des résultats
5. Application de méthodes avancées de data mining

Intérêt pédagogique

Le projet doit permettre aux étudiants de mettre œuvre et d'étendre leurs connaissances sur les points suivants :

- Découvrir l'outil Shiny de Rstudio pour le développement d'interface web interactive.
- Découvrir l'outil RGoogleAnalytics pour l'extraction des données d'audience des sites web
- Découvrir des outils de data Mining et de visualisation.
- Définir une interface riche avec Javascript, Html et R.
- Améliorer les compétences en programmation des bases de données

Outils mis en oeuvre : R, Javascript, Sqlite, Html, CSS

Encadrants : Lazhar Labiod

Références

1. <https://data-seo.fr/2015/08/21/ameliorez-donnees-google-analytics-r-shiny/>
2. <https://github.com/MarkEdmondson1234/googleAnalyticsR>
3. Package 'googleAnalyticsR' – CRAN.R-project.org
4. Emerson H., « Using Google Analytics with R », R-bloggers, August 2015 ; <https://www.r-bloggers.com/using-google-analytics-with-r/>
5. Granowitz A., « Using Google Analytics with R », September 2014 ; <https://developers.google.com/analytics/solutions/r-google-analytics>
6. Shah K., « How to extract Google Analytics data in R using RGoogleAnalytics », November 2014 ; <https://www.tatvic.com/blog/google-analytics-data-extraction-in-r/>
7. Wikipédia, « Audience d'un site web ».

Proposition de Projet Pluridisciplinaire pour la formation MAGE Paris Descartes

Information concernant l'encadrant

[Prénom Nom] Etablissement et poste : MILOUDI Fouad Chef de projet ERP

Email : fmiloudi@live.fr

Téléphone : 0685659505

Description générale du projet

Intitulé du projet : Application mobile avec géolocalisation pour transporter des objets volumineux entre particuliers et professionnels

Contexte :

Les smartphones font partie intégrante de notre société. Ils sont connectés au monde entier grâce aux technologies. L'une de ces technologies s'appelle la géolocalisation. Un très grand nombre d'applications intègrent la géolocalisation à leur catalogue de fonctionnalités proposées aux utilisateurs.

Objectifs :

- Aider les particuliers à transporter des objets volumineux plus rapidement et moins cher
- Aider les professionnels à proposer une prestation de service transport plus accessible

Réalisations attendues :

- Mise en place d'une application mobile avec la géolocalisation (android ou apple)

Niveau requis* : M1 et M2 (les deux)

** si M2, précisez s'il s'agit de la poursuite par une même équipe d'un sujet déjà abordé l'année précédente et indiquez l'équipe*

Description détaillée du projet

Description des tâches fonctionnelles :

- Etude du marché
- Rédaction d'un cahier des charges
- Rédaction d'une SFD/STD
- Cahier de test et manuel d'utilisation

Description des tâches techniques :

- Conception détaillée de l'application
- Déploiement de l'application

Compétences fonctionnelles et techniques requises (e.g. langage de programmation, outils de conception, bibliothèques) :

- JAVA ou autre
- Possibilité d'utiliser Android studio

Proposition de Projet Pluridisciplinaire pour la formation MIAGE Paris Descartes

Information concernant l'encadrant

Encadrant(s) : Séverine Affeldt et Lazhar Labiod, Maîtres de Conférences, Paris Descartes

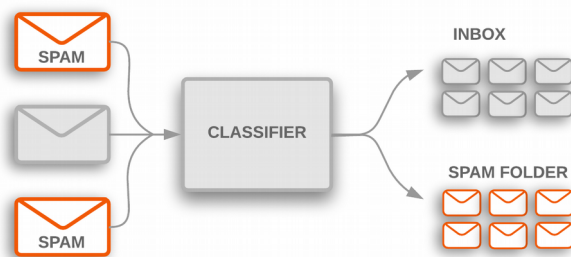
Email : severine.affeldt@parisdescartes.fr, lazhar.labiod@parisdescartes.fr

Description générale du projet

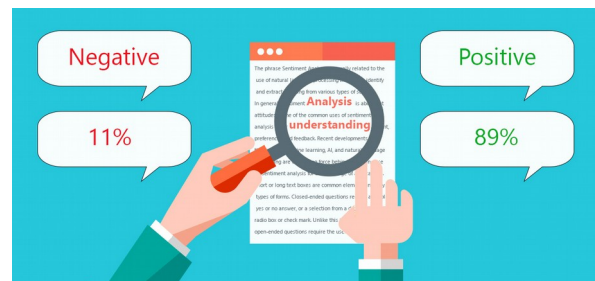
Intitulé du projet : Interface web pour la classification de textes par réseau de neurones

Contexte: La classification de textes est une application très courante dans de nombreux systèmes que nous rencontrons tous les jours. Elle intervient par exemple dans la gestion des emails (eg., email ou spam ? Fig. 1a), sur les forums de discussion (eg., identification automatique de contenu inapproprié) ou encore dans le domaine du marketing (eg., analyse de sentiments appliquée au Tweets, Fig. 1b).

Gestion d'emails – Fig. 1.a



Analyse de sentiments – Fig. 1.b



Ces outils, généralement basés sur des réseaux de neurones, dépendent de nombreux paramètres (eg., profondeur du réseau, largeur des couches du réseau). Ces paramètres sont parfois mal compris et utilisés, ce qui conduit à des classifications de mauvaise qualité.

Objectifs:

Ce projet a pour objectif l'implémentation d'une interface web pour l'aide au design et à l'utilisation d'un classifieur de textes qui se base sur un réseau profond. L'interface devra permettre à un utilisateur ayant peu ou pas d'expérience en Machine Learning de classer un ensemble de documents en un nombre de catégories prédéfinies.

Réalisations attendues (e.g. site web, application mobile) :

La première réalisation attendue pour ce projet est l'implémentation d'un classifieur de textes. Le groupe pourra pour cela utiliser les nombreuses ressources en lignes qui proposent beaucoup d'éléments de code permettant d'implémenter ce type de classification. La seconde réalisation est la mise au point d'une interface simple d'utilisation qui permettra l'exploitation du classifieur, c'est-à-dire son paramétrage, l'exploration des données en entrée et la visualisation des résultats de classement.

Niveau requis* : M1 ou M2

Description détaillée du projet

Description des tâches fonctionnelles :

- Plannification du projet
- Rédaction des documents de pilotage
- Rédaction de l'expression des besoins, des spécifications et analyse

Description des tâches techniques :

- Implémentation d'un classifieur de textes (de nombreux exemples d'implémentations existent déjà en ligne. Beaucoup d'exemples comportent des éléments des bibliothèques Python Tensorflow et Keras)
- Apprentissage sur l'ensemble des textes IMDb (Internet Movie Database; ce jeu de données bien connu contient les reviews de films. Beaucoup d'éléments de code sont à disposition en ligne pour exploiter facilement ces données)
- Identification des hyperparamètres les plus importants pour le réseaux de neurones
- Exploration des données et visualisation des résultats via l'interface

Compétences fonctionnelles et techniques associées (eg., langage de programmation, outils de conception, bibliothèques) :

- Python
 - Tensorflow et Keras
 - R Shiny
-