



Stages 2014-2015

MATHEMATIQUES/STATISTIQUES

Contact : Mme Lapedra, stage@isoftware.fr

ISOFT : UNE SOCIETE INNOVANTE

ISoft, éditeur de logiciels, est spécialisé dans **l'informatique décisionnelle et l'analyse de données**. Son expertise repose sur le traitement de **très gros volumes de données en temps réel**.

La société a 4 domaines d'activités :

- **Le Marketing et le E-Commerce**, avec une gamme d'outils de CRM analytiques et de Marketing Digital
- **La Monétique**, avec la lutte contre la fraude
- **Le Big data**
- **Les Sciences de la vie**, avec sa plateforme de bio-informatique

Depuis plus de 20 ans, ISoft, PME innovante, s'est imposée sur ses marchés (Data Management, Analyse de données...) et est leader sur le marché de la lutte contre la fraude bancaire. Sa clientèle est composée **exclusivement de Grands Comptes**.

ISoft a pour engagement d'apporter à ses grands donneurs d'ordres des **solutions stratégiques, innovantes et fiables** mettant en œuvre de nouvelles technologies. Pour ce faire, ISoft a une activité de Recherche importante. Société **en croissance**, disposant d'un environnement technique de premier plan, ISoft propose des **opportunités de carrière** au sein de ses activités .

Les stages, proposés tout au long de l'année et généralement d'une durée minimale de 4 mois, sont principalement destinés à des ingénieurs dans leurs dernières années d'études, qui désirent évoluer dans le domaine des nouvelles technologies, qui ont un goût prononcé pour l'innovation et souhaitent évoluer au sein d'une équipe de professionnels expérimentés et d'un environnement technique de qualité.

Vous trouverez ci-après nos offres de stages en

- **Mathématiques de la Décision, Data-Mining et Analyse de Données**

qui se répartissent selon les thématiques suivantes :

- **Big Data**
- **Modélisation risque de fraude**
- **Marketing digital**
- **Réseaux sociaux**



STAGES BIG DATA

Le **Big Data** s'accompagne du développement d'applications à visée analytique, qui traitent les données pour en tirer du sens. Elles portent sur des données quantitatives complexes avec des méthodes de calcul distribué. Nous proposons ci-dessous différents stages qui portent sur de l'interfaçage avec la technologie **Hadoop** et l'utilisation de langages de haut niveaux facilitant les calculs distribués.

Stage « Calculs distribués sur Hadoop »

La technologie **Hadoop** s'enrichit régulièrement de nouveaux modules tels que **YARN**, **ZooKeeper**, etc. facilitant sa mise en œuvre et destinés à étendre ses usages largement au-delà des applications « map-Reduce ».

L'objectif de ce stage sera de participer à la réalisation d'une application qui valide l'intégration de nos moteurs de calcul sur des environnements distribués de type **Hadoop**. Vous définirez l'architecture d'une application de traitement de données distribué mettant en œuvre **Hadoop** et dont vous implémenterez un prototype destiné à identifier les développements à réaliser sur les interfaces de nos moteurs de calcul.

Stage « intégration de langages appropriés aux calculs distribués »

Afin de simplifier l'écriture d'**applications distribuées** dans nos applications, nous souhaitons utiliser des langages de haut niveau, appropriés aux applications distribuées (**Julia**, **Python**, **Javascript**, ...).

Dans ce cadre, ce stage aura pour but la création d'une API d'extension pour différents langages afin de les intégrer dans nos architectures.

Stage « optimisation performance génération de code à la volée »

Afin d'améliorer les performances de nos applications distribuées, nous voulons étudier la génération de code à la volée du compilateur **LLVM** dans nos moteurs de calcul.

Dans le cadre du stage, vous serez amené à étudier l'utilisation du compilateur **LLVM** pour la génération de code à la volée, puis participer à son intégration de nos moteurs de calculs.

Technologies utilisées : C++ et les langages à intégrer.



STAGES MODELISATION RISQUE DE FRAUDE

Stage « Classification des comportements d'utilisation d'une carte bancaire »

L'usage de la carte bancaire s'est très fortement démocratisé ces dernières années. Au détriment du chèque, la carte bancaire est aujourd'hui utilisée pour effectuer la majeure partie des paiements. Les transactions réalisées portent unitairement peu d'information (date / heure, montant, type d'achat, pays dans lequel est réalisé l'achat et identifiant du terminal de paiement), cependant, l'analyse de l'historique des paiements d'une carte permet de décrire un ensemble important de comportements (fréquence d'utilisation, habitudes d'achat, utilisation uniquement pour retirer de l'argent dans un distributeur, utilisation le WE ...).

A partir de données transactionnelles, notre objectif est de déterminer le comportement d'utilisation d'une carte bancaire de chaque client. Une fois ce comportement modélisé, nous souhaitons identifier une modification anormale de son comportement sur un très court délai (utilisation frauduleuse de la carte à l'insu de son propriétaire).

A partir d'indicateurs de comportements d'utilisation d'une carte bancaire pré-existants, l'objectif de ce stage sera de déterminer une méthodologie de classification permettant d'obtenir des classes ou des groupes de classes stables dans le temps.

Stage « Modélisation avec des données censurées »

Pour la lutte contre la fraude à la carte bancaire, la mise en place de modèles statistiques ou de règles d'expert permet de détecter et d'éviter des transactions frauduleuses. Un modèle statistique efficace est un modèle qui évite le risque de fraude sans gêner les clients "légitimes".

Lors d'une phase de ré-estimation d'un modèle existant, nous devons prendre en compte le fait que ce dernier a pu refuser à tort des transactions non frauduleuses

Le stage visera à définir et à tester une stratégie de modélisation permettant de prendre en compte l'impact sur le comportement de fraude de modèles et règles pré-existants.

STAGES MARKETING DIGITAL

ISoft a développé une offre logicielle **CRM et Marketing digital** à destination des grands groupes. Son moteur de calcul, parmi les plus performants du marché, permet de proposer des fonctionnalités telles que la **personnalisation et le « Push » d'offres e-commerce** en temps réel. Les stages suivants visent à transposer les techniques de vente de la grande distribution au web en enrichissant les fonctionnalités de notre moteur.

Stage Moteur de Recommandations

Le stage consistera à réaliser la **modélisation** mathématique de **recommandations** aux clients, et/ou l'implémentation d'un moteur de recommandations qui pourra s'appliquer aussi bien sur la navigation web que sur les promotions offertes en magasins ou que sur des campagnes **marketing ciblées**. La personnalisation s'appuiera sur des données hétérogènes comme les données de navigation, les profils socio-économiques, les historiques d'achat, etc. Les techniques pressenties de modélisation s'appuieront sur des algorithmes statistiques simples et des règles d'associations.



STAGES RESEAUX SOCIAUX

Les réseaux sociaux fournissent et fourniront de plus en plus de données. Or, pour les entreprises il est essentiel de les suivre pour adapter leurs offres aux nouvelles demandes et aux avis des clients s'exprimant sur ces réseaux. Le stage consistera à étudier les différents réseaux sociaux et les données disponibles pour proposer ensuite des modèles d'analyse et des algorithmes adaptés tant au volume qu'à la nature de ces données.

Stage "L'analyse de données textuelles (ADT)"

Un grand nombre d'informations qualitatives relatives à la satisfaction clients transitent par les réseaux sociaux. Toutefois ces informations sont exprimées en langage naturel. Afin de valoriser ces informations, il est nécessaire de réaliser des analyses textuelles.

L'Analyse de Données Textuelles consiste à repérer les concepts présents dans un texte, leurs contextes d'utilisation et à en faire l'analyse. Elle fait appel à des algorithmes de traitement du langage naturel (phonétisation, analyse syntaxique, lemmatisation), et à des techniques statistiques (études des associations, des co-occurrences et d'analyses factorielles).

Le stagiaire aura à étudier différents outils d'ADT et à prototyper des algorithmes de phonétisation, de lemmatisation et d'analyse de données. Ces algorithmes seront testés sur des jeux de données réelles dans le domaine de la **satisfaction client** (données issues notamment de **blogs**).