

| | | |
|--------------------------------|---|-----------|
| SG053 6ICG093 | Analyse de données et Data Mining | S9 |
| Responsable : | Yves GUENIFFEY, Maître de Conférences Yves.Gueniffey@mines.inpl-nancy.fr | |
| Durée du module : | 42 heures | |
| Crédits ECTS : | 4 | |

Objectifs pédagogiques

Dans les années 70-80, le développement des ordinateurs a conduit au stockage d'informations dont la forme la plus classique était celle qui correspondait à des tableaux de données, généralement de grandes dimensions. Dans de nombreux domaines (géologie, météorologie, médecine, économie, marketing, contrôle de qualité, reconnaissance des formes...), l'analyse de données a permis de tirer parti de cette information pour la synthétiser, pour servir de base à un processus de décision, ou plus généralement, pour appréhender d'une certaine manière la nature des phénomènes sous-jacents aux données.

Depuis les années 90, la numérisation systématique de l'information fait que les organismes, publics ou privés, accumulent des masses considérables d'informations stockées dans des bases de données numériques, amorphes et dynamiques, données faites de chiffres, de textes, d'images, de sons, etc. Le Data Mining correspond à une « industrialisation » de l'analyse de données pour permettre une exploitation réelle du capital d'informations de l'entreprise : «extraire le minerai précieux de la gangue des données».

Contenu - Programme

Le programme porte sur les principales méthodes d'analyse de données et du Data Mining :

- Analyse en composantes principales
- Analyse des correspondances
- Analyse discriminante
- Classifications automatiques
- Discrimination et classification neuronales
- Segmentation

Leur mise en œuvre pose à l'utilisateur un certain nombre de questions dont les principales sont :

- Quels types de problèmes peut-on traiter ?
- Quelle méthode choisir ?
- Quelles données choisir ?
- Quels genres de résultats peut-on attendre ?
- Quelles en sont les limites ?
- Comment les mettre en œuvre ?

Un projet, réalisé en équipe, permettra à chaque élève, au-delà de l'apprentissage des techniques, d'apporter une réponse à ces questions et d'apprendre à utiliser un environnement logiciel moderne d'analyse de données (SAS, SAS Enterprise Miner et SPAD).

Des conférenciers invités viendront témoigner de leur expérience en matière d'analyse de données et de Data Mining.

Mode d'évaluation :

**Un travail personnel de synthèse écrit de 3h permettra une évaluation à mi parcours.
Le projet final donnera lieu à un rapport écrit et à une soutenance orale par équipe lors de la dernière séance.**

Références

- G. Saporta, *Probabilités, analyse des données et statistique*, Technip.
M. Tenenhaus, *Statistique, Méthodes pour décrire, expliquer et prévoir*, Dunod.
L. Lebart, A. Morineau et M. Piron, *Statistique exploratoire multidimensionnelle*, Dunod.
S. Tufféry, *Data Mining et statistique décisionnelle*, Technip.