



## **AVIS DE PRESENTATION DE THESE EN SOUTENANCE POUR L'OBTENTION DU DIPLOME NATIONAL DE DOCTEUR**

**Monsieur Nhu Khoa NGUYEN**

Présentera ses travaux intitulés :

**« Détection de tendances émergentes dans les articles d'actualité »**

Spécialité : Informatique et applications

**Le 27 février 2023 à 14h00**

Lieu :

**La Rochelle Université  
Pôle Communication, Multimédia et Réseaux  
Amphithéâtre Michel Crépeau  
44 Av. Albert Einstein  
17000 LA ROCHELLE**

Composition du jury :

**Mme BERTET Karell  
Mme BOROS Emanuela  
M. DELAHAUT Thierry  
M. DIAS Gaël  
M. DOUCET Antoine  
M. LEJEUNE Gaël  
M. MAUREL Fabrice  
Mme VILNAT Anne**

**Professeure, La Rochelle Université  
Research Scientist, Ecole Polytechnique Fédérale de Lausanne  
Professeur, Quantitative portfolio Manager, La Banque Postale  
Professeur, Université Caen Normandie  
Professeur, La Rochelle Université  
Maître de conférences, Sorbonne Université  
Maître de conférences HDR, Université Caen Normandie  
Professeure, Université Paris Saclay**

### **Résumé :**

Dans le domaine de la finance, l'information joue un rôle extrêmement important dans la prise de décisions en matière d'investissement. En effet, une meilleure connaissance du contexte peut conduire à l'élaboration d'approches plus appropriées quant à la manière d'investir et à la valeur de l'investissement. En outre, être capable d'identifier les thématiques émergentes fait partie intégrante de ce domaine, car ceci peut aider à prendre de l'avance sur les autres investisseurs, et donc à obtenir des avantages concurrentiels considérables. Pour identifier les thèmes susceptibles d'émerger à l'avenir, des sources telles que les rapports financiers annuels, les données des marchés boursiers ou encore les résumés des réunions de la direction sont examinées par des experts financiers professionnels. Des sources d'information fiables provenant d'éditeurs de presse réputés peuvent également être utilisées pour détecter les thèmes émergents. Contrairement aux médias sociaux, les articles de ces éditeurs jouissent d'une crédibilité et d'une qualité élevées. Ainsi, lorsqu'ils sont analysés en grande quantité, il est probable que l'on découvre des informations dormantes/cachées sur les tendances ou ce qui peut devenir des tendances futures. Cependant, en raison de la grande quantité d'informations générées chaque jour, il est devenu plus exigeant et difficile d'analyser les données manuellement tout en détectant les tendances au plus vite.

Notre recherche explore et analyse des données de différentes sources de qualité, telles que des résumés de publications scientifiques et un ensemble de données d'articles d'actualité fournis par Bloomberg, appelé Event-Driven Feed (EDF), afin d'expérimenter la détection des tendances émergentes. En raison de l'énorme quantité de données disponibles réparties sur de longues périodes de temps, elle encourage l'utilisation d'une approche contrastive pour mesurer la divergence entre le contexte environnant, passé et présent des mots et des phrases extraits, comparant ainsi la similarité entre les représentations vectorielles uniques de chaque intervalle pour découvrir des évolutions dans l'utilisation des termes qui peuvent conduire à la découverte d'une nouvelle tendance émergente. Les résultats expérimentaux révèlent que l'évaluation de l'évolution dans le temps du contexte des termes est susceptible de détecter les tendances critiques et les points d'émergence. On découvre également que l'évaluation de l'évolution du contexte sur une longue période est préférable à la simple comparaison des deux points les plus proches dans le temps.