

# Faith Lee

141 rue Cooper, Ottawa, Ontario, Canada

☎ 647-914-8045 | ✉ faithghlee@gmail.com | 🏠 randomstatsmusings.wordpress.com | 📷 faithghlee | 📺 faith-lee-6257b193/

## Sommaire

---

Je travaille actuellement en tant que scientifique de données dans une start-up à Ottawa qui développe un logiciel pour aider les auditeurs à analyser leur grand livre comptable. J'ai environ deux ans d'expérience dans le domaine de l'analyse de données. Je suis passionnée d'apprendre plus sur l'analyse de la survie et l'analyse sur les données longitudinales.

## Formations

---

### Université de Waterloo

MAÎTRISE EN MATHÉMATIQUES (BIOSTATISTIQUES), GPA : 4.00/4.00

Waterloo, Ontario

septembre 2016 - avril 2018

- Essai : Attenuation from intermittent measurement of time-dependent co-variates in multi-state models

### Université Memorial de Terre-Neuve

BACCALURÉAT SPÉCIALISÉ AVEC MAJEURE EN STATISTIQUES, GPA : 3.81/4.00

Saint-Jean de Terre-Neuve

septembre 2013 - août 2016

- Thèse : Comparison of logistic and survival regression models

### Polytechnique Ngee Ann

DIPLOME EN SCIENCE BIOMÉDICALE, GPA : 3.67/4.00

Singapour

avril 2010 - mai 2013

- Essai : Characterization of NDM-1 and IMP genes in antibiotic resistant bacteria

## Expériences professionnelles

---

### MindBridge Analytics Inc.

SCIENTIFIQUE DE DONNÉES

Ottawa, Ontario

juin 2018 - maintenant

- Recherché les analyses possibles basées sur l'industrie et les données du client (dans les domaines d'assurance et de prêts) selon l'objectif de l'utilisateur final. Régression des moindres carrés ordinaires, modèle linéaire général, K-moyennes sont parmi des méthodes populaire pour analyser les données.
- Nettoyé des grosses ensembles de données en utilisant les commandes Unix (> 100 million de rangées)
- Créé les visualisations de données pour communiquer avec les clients à propos de leurs données (par exemple, les graphiques de carte thermique, 'forest plots', 'lattice plots')
- Écrit du code de niveau production en Python comme objectif d'intégrer dans le backend
- Fait des recherches dans différents domaines comme vision numérique et traitement automatique du langage naturel (NLP). Par exemple, réduire le bruit dans les images (image d'une facture), convertir une image vers du texte, ensuite analyser ce texte-là avec l'aide du word embedding (plongement lexical). Tesseract, Azure Optical Character Recognition (OCR) API, OpenCV, Keras/Tensorflow, spaCY sont parmi les outils que j'utilise
- Aidé dans la prise de décisions en utilisant les méthodologies orientées données
- Technologies utilisées : Python, R, Azure Databricks, Azure OCR, CircleCI, JIRA, Docker

### Institut des sciences évaluatives cliniques

ANALYSTE (STAGIAIRE)

Toronto, Ontario

mai 2017 - décembre 2017

- Tiré les informations importantes sur la santé de la population par l'usage des bases de données provinciales
- Traité des données dans une forme souhaitable pour analyse et présentation (par exemple, fusionnage et nettoyage de données)
- Utilisé les modèles statistiques pour approfondir notre connaissance sur ces données, par exemple l'usage d'un modèle multi-état pour examiner les femmes qui font au dépistage du cancer de cervix, et binomiale négative équation d'estimation généralisée (EEG) pour mieux comprendre la vraisemblance des gens à prendre ce sondage
- Contribué à l'écriture de manuscrit pour l'objectif de publication, validé les résultats qui viennent de la modélisation de données et créé les visualisations pour les publications)
- Technologies utilisées : SAS, R

### Département de statistique et de science actuarielle, Université de Waterloo

AUXILIAIRE D'ENSEIGNEMENT

Waterloo, Ontario

septembre 2016 - avril 2018

- Auxiliaire d'enseignement pour les cours basiques en statistique
- Mes responsabilités étaient de corriger les examens et les devoirs, de superviser les examens, de tenir des heures de consultation et de répondre aux questions en-ligne

## Publications

---

## Immigration and adherence to cervical cancer screening : a province-wide longitudinal matched cohort study using multistate transitional models

Lee F., Sutradhar, R. and Paszat, L.

Journal of Obstetrics and  
Gynaecology Canada

## Higher risk of gastric cancer among immigrants to Ontario : a population-based matched cohort study with over 2 million individuals

Sutradhar, R. Asidiyana N., Lee F., et. al.

Gastric Cancer

## Use of patient reported outcomes in regional cancer centres : a retrospective study

Barbera, L., Lee, F and Sutradhar, R.

Canadian Medical Association  
Journal

## Bourses d'études

---

### UNIVERSITÉ DE WATERLOO

International Masters Student Award  
Statistics and Actuarial Science Chair's Award

Fall 2016, Winter 2017, Winter 2018  
Winter 2017, Winter 2018

### UNIVERSITÉ MEMORIAL DE TERRE-NEUVE

Faculty of Science Dean's List  
MUN Faculty of Science Summer Undergraduate Research Award  
The Williams Science Scholarship  
Dr. B.K. Kim Scholarship in Statistics  
International Undergraduate Scholarship

2013-2014, 2015-2016  
2015, 2016  
2014-2015  
2014-2015  
2013-2014

## Conférences

---

### Science Atlantique mathématiques, statistique et informatique conférence, 2017

Fredericton, Nouveau-Brunswick

PRESENTATRICE

octobre 2017

- Titre : Multi-state modelling for examining adherence to cervical cancer, basé sur recherche pendant mon stage

### Science Atlantique mathématiques, statistique et informatique conférence, 2015

Wolfville, Nouvelle-Écosse

PRESENTATRICE - 3ÈME PRIX SOUS LA CATÉGORIE BEST ORAL PRESENTATION (STATISTICS)

octobre 2015

- Titre : Comparison of the logistic regression and Cox regression model, basé sur recherche d'été

## Développement professionnel

---

### COURS EN LIGNE/ MOOC

Bayesian Methods for Machine Learning sur Coursera  
Convolutional Neural Networks sur Coursera  
Improving Deep Neural Networks : Hyperparameter tuning, Regularization and Optimization sur Coursera  
Neural Networks and Deep Learning sur Coursera  
Machine Learning sur Coursera  
Practical Machine Learning sur Coursera  
Java Programming : Arrays, Lists, and Structured Data sur Coursera  
Java Programming : Solving Problems with Software sur Coursera  
R Programming sur Coursera

2019  
2018  
2018  
2018  
2017  
2017  
2019  
2019  
2017

### CERTIFICATIONS

Diplôme d'études en langue française (DELF) B2  
Diplôme d'études en langue française (DELF) B1  
1Z0-051 Oracle Database 11g : SQL Fundamentals I  
SAS Certified Statistical Business Analyst  
SAS Certified Advanced Programmer  
SAS Certified Based Programmer  
Society of Actuaries Exam P (7/10)

2019  
2019  
2017  
2016  
2015  
2015  
2015

## Compétences Spécifiques

---

**Langages de programmation** Python, R, SAS, LaTeX

**Gestion de version** Git

**Langues** Anglais (bilingue), mandarin chinois (bilingue), français (intermédiaire avancée)