TEXT ANALYSIS CERTIFICATE

Program Description

In today's AI-driven landscape, processing and analyzing textual data is increasingly critical for understanding customer sentiment, market trends, and other key business insights. Natural language processing (NLP) techniques have become essential tools for transforming raw text into meaningful data that underpins many of today's AI applications.

This certificate program is designed to equip you with foundational skills in NLP, with a focus on text preprocessing, summarization, visualization, and sentiment analysis. In the first course, you will clean and manipulate text data using regular expressions, preprocess complex textual information, and address common challenges in messy datasets. You will also have the opportunity to explore advanced text preprocessing techniques such as stemming and tokenization, which are essential for preparing text for further analysis. In the second course, you will develop the ability to summarize and visualize text distributions across documents, leveraging tools like word clouds and document-term matrices to uncover patterns and trends. Finally, the third course introduces you to sentiment analysis, where you will quantify and interpret emotions in text and compare sentiment across documents and over time.

By the end of this program, you will have the practical knowledge needed to preprocess and analyze textual data, giving you a valuable edge in data science, AI engineering, or any field that requires a deep understanding of textual information.

To succeed in this program, you should have a foundation in R programming. If you do not have this experience, start with the Data Science Essentials certificate program.

The courses in this certificate program are required to be completed in the order that they appear.

Key Takeaways

- Clean and preprocess the textual data contained within a set of documents in preparation for sentiment analysis
- Summarize and visualize the distribution of words within a single document (univariate) and across multiple documents (multivariate)
- Compare word distributions across documents and over time
- Use three different sentiment analysis lexicons (AFINN, Bing, and NRC) to quantify and interpret sentiments associated with words, sentences, and paragraphs
- Compare sentiments across documents and over time

What You'll Earn

- Text Analysis Certificate from Cornell's Ann S. Bowers College of Computing and Information Science
- 48 Professional Development Hours (4.8 CEUs)

Who Should Enroll

- Data scientists
- Computer scientists
- Analysts
- User behavior and UX teams
- Researchers
- Social scientists

Courses

Code	Title	Hours
eCornell CIS611	Mastering NLP Fundamentals	0
eCornell CIS612	Exploring Summarization and Visualization	0
eCornell CIS613	Transforming Text to Numeric Sentiments	0