Andres Cremisini

Data Scientist

Miami, FL

1.9546125378

acremisini@gmail.com

Cover Letter

I am a data scientist with a specialized skill set in natural language processing and a varied yet complementary intellectual and occupational background. Before beginning my career as a data scientist, I studied and worked as a composer. In 2013, I graduated with a M.A. in Music Composition from U.C. Berkeley with a focus on Improvisation and Music Cognition. I have been a teacher for about a decade, teaching privately, in the classroom and as a mentor for Senior Undergraduate students in Computer Science at Florida International University (FIU). The unifying theme in my professional career is a strong fascination for patterns and how humans work on both an individual and societal level. To introduce myself professionally, I will briefly describe some of my projects in backwards chronological order.

In the Summer of 2019 I worked as a Data Science Fellow at the Bureau of Labor Statistics (BLS), the principal fact finding agency of the United States Government. I was there as part of *Coding It Forward*, a competitive fellowship program (~5% acceptance rate) that brings students from around the U.S. to work in Civic Tech in Washington, D.C for 10 weeks. At BLS, I built a language classifier to automate a very large and expensive language classification task. The classifier performs at > 97% accuracy, and lays the groundwork to automate a tedious process that costs millions of dollars and improves the quality of BLS statistics, which are widely used by Government and Business to make decisions that affect millions of people. In addition to the classifier, I built an intuitive and interactive dashboard that allows non-technical economists to train and analyze different machine learning models.

As a graduate student in data science, I conduct original NLP research to automatically detect disinformation in long-form text, fully funded by the Cognition, Narrative and Culture Laboratory at Florida International University. I have recently been published by Springer as first author at the 12th annual SBP-BRiMS conference in Washington, D.C., an important conference for Computational Social Science, for part of this work. I am developing software to solve the difficult task of Cross Document Event Coreference (CDEC), where I have made significant strides and am in the process of both writing a book chapter on CDEC to be published by Springer as well as a technical paper as first author describing the CDEC work.

A personal project during my Masters involved implementing a state of the art Deep Learning language model called a Siamese-LSTM, in pytorch. It is the first step in a clustering system designed to cluster many

natural language comments, allowing decision makers to visualize the semantic clusters in a large corpus of comments. The motivating use case for this system is decision makers in government responding to constituent comments, though it could be used by any organization wishing to understand a large volume of text.

I am motivated by challenging, nuanced technical problems that require understanding and empathizing with some human process, hopefully while working with and learning from good colleagues.

Thank you for your time. I look forward to speaking!

Attentively,
Andres Cremisini