

UNDERGRADUATE MATH SEMINAR

The next seminar of the winter term will be

DATE: **THURSDAY, April 23**

Time & **12:30 – Refreshments in Bailey 204**

Location: **12:50 – 1:45 Seminar in Bailey 207**



Professor Vincent
Martinez

In this seminar, **Professor Vincent Martinez** from the Department of Mathematics at CUNY Hunter College and CUNY Graduate will present the following talk.

Title: **Why Can We Predict the Weather and Why Is It Hard To Do?**

Abstract: Predicting the weather is an old problem and it's still unsolved in spite of having access to weekly forecasts on our phones. We've gotten very good at it over the years though and have developed more principled methods that do away with things like star gazing, watching animals, and folklore. As our technology and understanding advanced, we've been able to collect more data about the atmosphere and have a better physical understanding of the physical mechanisms involved. The modern method is to incorporate weather data into partial differential equations that model the weather. This practice is often referred to as *data assimilation*. This talk will be about the mathematics of data assimilation; we will present a mathematical reason for why the modern method is so successful but also indicate why predicting the weather is still a difficult thing to do.

Pieces from Thesis: Talha Khan

*Talha wrote a joint Math-CS senior thesis this past fall and winter under the guidance of **Professors Roger Hoerl and Nick Webb**.*

Over the course of this past Fall and Winter, I had the great pleasure of working with Professors Roger Hoerl (Statistics) and Nick Webb (Computer Science) on a joint thesis project titled "Algorithmic Evaluation of the Soccer Transfer Market." This was a joint thesis between the Mathematics and Computer Science departments, which is somewhat uncommon as the Computer Science department is fairly specific in terms of senior capstone requirements. I had to submit a formal proposal in the Spring of my junior year outlining what the project would entail before it could be approved. I remember meeting with the two professors over Zoom and sharing my ideas for what I wanted to accomplish, as well as the tools that would be required.

The thesis focused on the European soccer transfer market. Unlike American sports, European leagues do not operate on a draft system; instead, clubs pay negotiated fees to other clubs to acquire their players. In the summer of 2009, for example, Spanish club Real Madrid paid Manchester United a then-world-record £80 million to sign Cristiano Ronaldo, a transfer that proved to be an exceptional investment as Ronaldo went on to cement his legacy as one of the greatest players of all time. On many other occasions, however, players have performed far below what their transfer fee could justify. The goal of my project was to determine whether a player's transfer fee could be predicted algorithmically from their measurable attributes, and whether certain players were systematically over- or undervalued by the market.

The first term was spent almost exclusively on data collection. No clean, publicly available dataset combined historical transfer fees with comprehensive player statistics, so I assembled one from

Continued on the next page.

scratch. Using an R package that interfaces with the well-known transfer database Transfermarkt, I collected over 1,200 Premier League transfer records spanning ten seasons. For each player, I matched their transfer record to their FIFA video game ratings, a widely used and validated proxy for real-world ability, capturing attributes such as overall rating, potential, pace, passing, dribbling, and physicality. The data cleaning and merging process was far more time-consuming than I had anticipated, but it produced a novel dataset that does not exist anywhere else.

With the data in hand, we trained and evaluated nine predictive models ranging from simple linear regression to gradient-boosted trees and neural networks, predicting the log-transformed transfer fee. Transfer fees are logged because they grow exponentially and expressing them on the log scale makes the modeling problem much more tractable. Perhaps the most surprising finding was that ordinary linear regression outperformed every more complex model on the held-out test set, achieving an R^2 of 0.59. This turned out to make intuitive sense: once fees are on the log scale, the relationship between player ability and fee is largely linear, and the residual unpredictability reflects factors that are difficult to quantify, such as negotiation dynamics, contract length, agent influence, and commercial considerations.

The dominant predictors were a player's overall FIFA rating and their potential rating, consistent with the idea that clubs pay not just for what a player can do now, but for what they expect them to become. Positional effects were also significant, with defenders and midfielders commanding systematically lower fees than forwards of comparable ability, reflecting both the lower commercial value of defensive players and the difficulty of quantifying their contributions. Younger players were substantially overvalued relative to older ones, and 75% of transfers in any given season involved players younger than 27. Transfer fee and age followed a parabolic relationship.

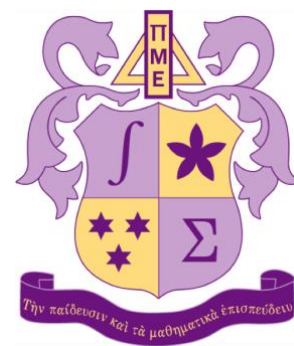
This project taught me many things: that rigorous data collection is half the battle, that interpretable models can outperform complex ones, that it is far more interesting to use data to tell a narrative than to simply crunch numbers, and that a senior thesis should be treated as both a learning experience and an opportunity to work on something you are genuinely passionate about, because not everything will go according to plan, and that is perfectly fine.

Pi Mu Epsilon. Application Deadline Noon, Thursday, April 23

The Union College chapter of Pi Mu Epsilon (PME), a national undergraduate math honors society, is accepting applications for membership. To be considered for membership into Union's chapter of PME, a student must have

- taken at least two math courses at the 200-level or above;
- a minimum campus-wide GPA of 3.0 and a minimum GPA of 3.0 in all math courses that can be used to fulfill a math major;
- a demonstrated extracurricular involvement in mathematics while at Union;
- an interest in joining PME, demonstrated by writing a few paragraphs/mini-essay in an email to **Professor Paul Friedman** (friedmap@union.edu) explaining your interest, your qualifications, and how you have supported and will continue to support the purpose and goals of PME. This email must be received by **noon on Thursday, April 23**.

For more information about PME, go to the website pme-math.org.



CALCULUS HELP CENTER!

Sunday, Tuesday, and Thursday: 7:30 – 10:00pm, Sorum House Seminar Room