Logo, company name

Description automatically generated

**George Mason University**

01/29/2023

**Fall  
2023**

Darvik Kunal Banda

Erick A Torres

Joe B Brock

Sagar D Goswami

Telina Andrianaivo

**DAEN 690  
Project Report**

**Predicting CEC Graduate Course Demand**



Table of Contents

[Abstract 6](#_Toc125926368)

[1 Introduction 7](#_Toc125926369)

[1.1 Background 7](#_Toc125926370)

[1.2 Problem Space 7](#_Toc125926371)

[1.3 Research 8](#_Toc125926372)

[1.3.1 Forecasting approaches in a higher education setting (Hassan Bousnguar, 2022) 8](#_Toc125926373)

[1.4 Solution Space 8](#_Toc125926374)

[1.5 Project Objectives 9](#_Toc125926375)

[1.6 Primary User Stories: 9](#_Toc125926376)

[1.7 Product Vision - Sample scenarios (why would someone want to use this) 10](#_Toc125926377)

[Scenario #1 10](#_Toc125926378)

[Scenario #2 10](#_Toc125926379)

[Scenario #3 10](#_Toc125926380)

[1.8 Definition of Terms: 10](#_Toc125926381)

[2 Data Acquisition 12](#_Toc125926382)

[2.1 Overview: 12](#_Toc125926383)

[2.2 Field Descriptions: 12](#_Toc125926384)

[2.3 Data Context: 12](#_Toc125926385)

[2.4 Data Conditioning 12](#_Toc125926386)

[2.5 Data Quality Assessment: 13](#_Toc125926387)

[2.6 Other Data Sources 13](#_Toc125926388)

[3 Analytics and Algorithms 14](#_Toc125926389)

[4 Visualization 15](#_Toc125926390)

[5 Findings 16](#_Toc125926391)

[6 Summary 17](#_Toc125926392)

[7 Future Work 18](#_Toc125926393)

[8 Appendix A: Code References 19](#_Toc125926394)

[9 Appendix B: Risk Section 20](#_Toc125926395)

[9.1 Sprint 1 Risks 20](#_Toc125926396)

[9.2 Sprint 2 Risks 20](#_Toc125926397)

[9.3 Sprint 3 Risks 20](#_Toc125926398)

[9.4 Sprint 4 Risks 20](#_Toc125926399)

[9.5 Sprint 5 Risks 20](#_Toc125926400)

[10 Appendix C: Agile Development 21](#_Toc125926401)

[10.1 Scrum Methodology 21](#_Toc125926402)

[10.2 Sprint 1 Analysis 21](#_Toc125926403)

[10.3 Sprint 2 Analysis 21](#_Toc125926404)

[10.4 Sprint 3 Analysis 22](#_Toc125926405)

[10.5 Sprint 4 Analysis 22](#_Toc125926406)

[10.6 Sprint 5 Analysis 22](#_Toc125926407)

[11 References 23](#_Toc125926408)

Table of Figures

**No table of figures entries found.**

Table of Tables

**No table of figures entries found.**

# Abstract

*[Write* ***one*** *paragraph of* ***no more than 300 words*** *that summarizes your project. Here are the typical kinds of information found in most abstracts which you should use as an outline as you develop your abstract.*

1. *The* ***context*** *or* ***background information*** *for your research; the* ***general topic*** *under study; the* ***specific topic*** *of your research.*
2. *The* ***central questions*** *or* ***statement of the problem*** *your research addresses.*
3. ***What’s already known*** *about this question, what* ***previous research*** *was conducted or shown.*
4. *The* ***main reason(s)****, the* ***exigency****, the* ***rationale****, the* ***goals*** *for your research — why is it important to address these questions? Are you, for example, examining a new topic? Why is that topic worth examining? Are you filling a gap in previous research? Applying new methods to take a fresh look at existing ideas or data? Resolving a dispute within the literature in your field?*
5. *Your* ***research*** *and/or* ***analytical methods****.*
6. *Your main* ***findings****,* ***results****, or* ***arguments****.*
7. *The* ***significance*** *or* ***implications*** *of your findings or arguments.*

*Your abstract should be intelligible on its own, without a reader’s having to read your entire paper.*

***NOTE: THE DRAFT ABSTRACT IS WRITTEN IN WEEK 12 OF THE PROJECT AND THE FINAL ABSTRACT COMPLETED IN WEEK 13 OF THE PROJECT****.]*

# Introduction

## Background

* *CEC is explanding at a rapid rapid pace*
* *Lots of New Enroillment*
* *Lots of new courses*
* *Pandemic had disrupted demand for Online/In-person classes*
* *Lots of students are being waitlisted*
* *Causes the students to defer graduation*
* *International Students are limited by what courses they can take*
* *Demand for Online Courses is more than in-person*
* *It is very difficult to add new sections in a short time frame*
* *Knowing about the demand in advance will enable proper resource allocation.*

George Mason University (GMU) is growing at a rapid pace, receiving roughly 8.41% more applications annually (Data USA). Moreover, lots of new courses are being introduced every year, which highlights the increasing reputation of the George Mason University in delivering to their mission statement. Perhaps with this incremental increase in student enrollments, it is becoming more difficult to manage course enrollments by the day. Adding to it, the 2019 pandemic has only made it more difficult with the introduction of online modality of education, and addition of many online courses that were only available in-person. This has resulted in a lot of students being waitlisted for their preferred courses. Students have to defer their graduation due to unavailability of their preferred courses, whereas international students have to take other courses to maintain their Full Course Load. Besides, a lot of students have their own preferences for online/in-person courses, which further adds to uncertainty in determining proper number of sections. These changes in the section are very difficult and oftentimes impossible to implement, as the resource allocation for these changes takes a lot of time. Knowing about the enrollment demand can provide enough time to the Enrollment Management at GMU to dedicate and allocate resources to satisfy course demand.

## Problem Space

* Focusing on CEC initially to understand the efficacy of these techniques for the problem at hand
* Analyze the Student Enrollment trends to predict the demand of courses in future

For this project, only CEC courses are focused, as it gives a good understanding and feedback on the efficacy of using Data Analytics and Machine Learning techniques in predicting the course demand. This purpose of this project would be to explore different algorithms and techniques and provide a foundation for a service that can be automated and further scaled to more departments and courses.

## Research

A lot of research indicates a huge scope for Machine Learning and Data Analytics techniques for effective Enrollment Management. These tools can help in Enrollment Retention, Course Suggestions, predicting course demand, and many more.

### Forecasting approaches in a higher education setting (Hassan Bousnguar, 2022)

This paper published in the Education and Information Technologies journal shows various algorithms and Educational Data Mining (EDM) principles to predict the demand for certain courses at the Ibn Zohr University, Morocco. They tested the performance of algorithms like Long Short-term Memory (LSTM) model, Auto-Regressive Integrated Moving Averages (ARIMA) and Exponential Smoothing model, as well as Fuzzy Time Series.

It was determined from this report, that for their data, Fuzzy Time-Series provided really good results compared to other algorithms. However, they noted that the performance of these models is greatly dependent on the data supplied. Thus, it becomes very important to test multiple algorithms for different scenarios. Moreover, further post-analysis should be performed to test it against any edge cases.

## Solution Space

The outcome of our system will allow GMU Enrollment Management, and those interested in their mission, to gain a better understanding of what classes will need to be offered. The outcome will provide specific courses to provide, a specific amount of sections to offer, and the number of chairs available in each section. This tool will allow GMU Enrollment Management to effectively address waitlist issues that they are currently facing. With accurate information being given to Enrollment Management, they can take effective action to ensure all students are getting the classes they need to graduate.

## Project Objectives

Providing a Course Management Recommendation Tool that can help with predicting course demand can offer many benefits to the team members, GMU CEC Graduate Student Population, and Staff. Regarding the team, we will have applied and grown skills in Machine Learning, Predictive Analytics, and Data Engineering. The team will have also gained insight into project management by applying the Agile Development Methodology to efficiently complete a large task. GMU CEC Business Intelligence staff will be more effective at recommending which courses to offer and how many sections. This type of knowledge give them foresight to know how much staff is required and save time guessing at how many classes are required in the future. With the CEC Business Intelligence staff having more insight into how to effectively, that will lead students to less frustration with their available classes and higher rates of graduating on-time. The new knowledge and insights gained from this project show how diverse backgrounds coming together to solve problems can achieve change on a large scale.

## Primary User Stories:

Based on the user context and value proposition, we developed the following primary user story to guide our project:

* As a user, I would like to get the forecasts of the course demand 2 years in the future.
* As a user, I would like to minimize the waitlist for the courses, to maintain the proper student enrollment.

## Product Vision - Sample scenarios (why would someone want to use this)

Our product will help prevent students from graduating at later date by predicting which classes are the most prone to waitlisting. This is beneficial for educational organizations because high graduation rates are appealing to parents and teachers.

### Scenario #1

The Office of Provost would benefit from minimized waitlists. Less courses available mean less students being able to take a class. Frustrated students create a more negative class experience, biasing students against the class. By reducing the waitlist, more students can participate in the classes thus generating more money to the university.

### Scenario #2

Waitlists can deter academic progress for professors, since adding students means allocating more resources. If individual students have issues with a class, their specific issues may take time to properly resolve. Professors who are unable to handle extra students or students that are likely to drop will have a harder time giving students the best quality education. Our tool will not fix the organization issues that professors may have, but it will provide professors better predict volatile waitlisting rates for a class.

### Scenario #3

Parents benefit knowing that their students will not be forced to apply for another class. Going to college is an expensive endeavor for schools who are retaining students and parents who are paying for the classes. Students who are waitlisted for classes will be in a worse position to graduate, and late graduations negatively affect job outlooks. Parents should be comfortable knowing that their adult children won’t pass their graduation date because their classes weren’t available when they needed them.

## Definition of Terms:

Deep Learning: Analysis utilizing Neural Networks and Processing

Enrollment Management: Management regarding Student Enrollment

Linear Programming: An Operations Research technique that attempt to identify minimized/maximized solutions for a given resource-based problem.

Machine Learning: Utilization of algorithms and models to identify and act on patterns of data without outside input.

Predictive Analytics: Utilization of algorithms and models to predict whether a provided event will occur or not.

CEC: College of Engineering and Computing

# Data Acquisition

## Overview:

*[Provide a descriptive overview of your datasets]*

## Field Descriptions:

*[Described your dataset field. Make sure you study the example below and you will more than likely expand these fields:*

* *URL (Type: string) – The web address or Universal Resource Locator for the webpage that contained the news article. This includes the protocol (http or https), host name, and subdomain. Some URLs also include parameters (text following ‘?’) or named anchors (text following a ‘#’). Each URL can only be present once in the database, even if the webpage is not static over time.*
* *Title (Type: string) – The title of the news article as parsed by the Newspaper 3K module. This field may be null (~150 articles in our dataset do not have titles).*
* *Authors (Type: string) –The authors of the news article as parsed by the Newspaper 3K module. This field may be null (~23,000 articles do not have authors) and articles with multiple authors have their names joined with a comma into a single string. This field may also pick up descriptions of the author, including their titles and background.*
* *Publication Date (Type: datetime) – The article publication date and time as parsed by the Newspaper 3K module. The datetime is displayed in ISO 8601 format (YYYY-MM-DD Thh:mm:ss+offset). Publish dates without specified times are assumed to be published at midnight. Publication dates with time information, but without a timezone listing, are assumed to be in Eastern Standard Time. This field is not allowed to be null.*
* *Text (Type: string) – The text of an article as parsed by Newspaper 3K. This field may be null (~8,000 articles do not have text) as some news stories are delivered as only video, audio, or a picture. The mean word count for text is 538.9 across all news sources.*
* *Tags (Type: string) – Article tags as determined by Newspaper 3K. These appear to be important (rare or “topicy”) words taken from the article text, not meta tags contained in the article’s HTML. Multiple tags are concatenated with a comma into a single string.]*

## Data Context:

*[Provide a description of the data context.]*

## Data Conditioning

*[Describe the data conditioning required for each data set.]*

## Data Quality Assessment:

*[At a minimum you must assess your data sets with the following attributes:*

* *Completeness:*
* *Uniqueness:*
* *Accuracy:*
* *Atomicity:*
* *Conformity:*
* *Overall Quality:]*

## Other Data Sources

*[If you are considered other data sources, however, you decided not to use these sources provide some reason why they were not utilized.**]*

# Analytics and Algorithms

*[Provide detailed descriptions of the analytics you used and how this related to your project analyses.]*

# Visualization

*[Describe and show findings and results based on a mix of figures and descriptive text. If you have video, it will be limited to presentation, however, it can also be reference as media file in your Blackboard file exchange.]*

# Findings

*[Summarize your findings and results for the reader.*

# Summary

*[Summarize your overall results for the reader. What did you discover, prove, disprove, etc.]*

# Future Work

*[Critical section! Propose future work or next step(s) for this project.]*

# Appendix A: Code References

*[Provide a GitHub Link and the README.MD content. Do not just provide a link to the GitHub repository but provide a narrative paragraph which introduces the project. This section should mirror the look and feel of a well-documented professional GitHub site.]*

# Appendix B: Risk Section

## Sprint 1 Risks

*[Include the risk table associated with the Sprint. Below the risk table provide a narrative description of how the risks and mitigation plans were identified, what the team got correct, what the team could have done differently, how accurate was the team in identifying the risks, did the team encounter any unanticipated risks, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Risk Name*** | ***Description*** | ***Probability*** | ***Impact*** | ***Mitigation*** |
| *Miscommunication* | *The team gets confused as to what is expected* | *Medium* | *High* | *Ensure all see the YouTrack and monitor WhatsApp* |
| *Commitment* | *Some members have more time than others.* | *Low* | *Medium* | *Balance workload depending on availability* |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

## Sprint 2 Risks

*[Include the risk table associated with the Sprint. Below the risk table provide a narrative description of how the risks and mitigation plans were identified, what the team got correct, what the team could have done differently, how accurate was the team in identifying the risks, did the team encounter any unanticipated risks, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 3 Risks

*[Include the risk table associated with the Sprint. Below the risk table provide a narrative description of how the risks and mitigation plans were identified, what the team got correct, what the team could have done differently, how accurate was the team in identifying the risks, did the team encounter any unanticipated risks, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 4 Risks

*[Include the risk table associated with the Sprint. Below the risk table provide a narrative description of how the risks and mitigation plans were identified, what the team got correct, what the team could have done differently, how accurate was the team in identifying the risks, did the team encounter any unanticipated risks, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 5 Risks

*[Include the risk table associated with the Sprint. Below the risk table provide a narrative description of how the risks and mitigation plans were identified, what the team got correct, what the team could have done differently, how accurate was the team in identifying the risks, did the team encounter any unanticipated risks, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

# Appendix C: Agile Development

## Scrum Methodology

*[Provide a narrative of the team efforts in using a scrum methodology for a data analytics engineering project. Describe how easy/difficult was it to adapt to the Scrum methodology. Did the team conduct a daily scrum? If not, how often did the team conduct a scrum. Describe how easy/difficult it was to use the YouTrack tool to manage the project. Don’t be limited to just these questions. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 1 Analysis

Sprint 1 was difficult. Our team had lost two of its members during the first week, but it was productive. We met with the Office of Provost to clarify our functional requirements and gain insight to what we were developing. Sagar Goswami was adopted as the Scrum Master and Erick Torres was adopted as Product Owner. The team did well on identifying some of the initial technical user stories, with Joe Brock handling potential Third-Party API Services and API Management. More work needs to be done in defining the stories listed, but this will likely be resolved when we receive the data provided.

*[Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 2 Analysis

*[Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 3 Analysis

*[Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 4 Analysis

*[Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 5 Analysis

*[Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

# References

Data USA. (n.d.). *George Mason University.* Data USA. Retrieved from https://datausa.io/profile/university/george-mason-university/#:~:text=George%20Mason%20University%20received%2021%2C198%20undergraduate%20applications%20in,accepted%20for%20enrollment%2C%20representing%20a%2089.2%25%20acceptance%20rate.

Hassan Bousnguar, L. N. (2022, August). Forecasting approaches in a higher education setting. *Education and Information Technologies*, 27. Retrieved from https://link.springer.com/article/10.1007/s10639-021-10684-z

*[INSTRUCTIONS: The References section of this document makes use of the Microsoft Word References feature to insert research citations by recording them directly into the document. All citations are to follow the IEEE citation format. Use the Bibliography drop down to have Microsoft Word dynamically create your Works Cited section here in IEEE citation format.*

*To learn more about the IEEE Citation guidelines click on the document links below.*

* [IEEE-Reference-Guide.pdf](https://ieeeauthorcenter.ieee.org/wp-content/uploads/IEEE-Reference-Guide.pdf)
* [IEEE Citation Guidelines2.doc (ieee-dataport.org)](https://ieee-dataport.org/sites/default/files/analysis/27/IEEE%20Citation%20Guidelines.pdf)

*Delete this red text section after you have read and understood the instructions.]*