



MATES Computer Science

Senior Capstone Project Bi-Weekly Progress Report

Project Title	MorningBread
Team Members	Max Bradshaw, Jake Dorick
Dates Covered by Report	February 26th-March 8th, 2024
Link to Github	https://github.com/maxbshaw17/MorningBread

1. **Summary of Project** (Provide a one paragraph summary of your project. You can largely copy/paste this from one progress report to the next, unless there are significant changes.)

To provide a personalized morning financial report to our users by utilizing a web scraping tool for financial website articles to find the general topics of the articles. In doing so, we are creating and designing a website to give the public a general grasp of financial news for the day, over the course of the day with hourly, highly summarized news headlines/stock tickers. An account system will allow users to follow certain companies and manage tickers.

2. **Summary of Progress this Period** (Provide a high-level, one-paragraph overview of what was accomplished this progress period collectively by the team.)

Jake-

The majority of progress for the frontend development came from the learning of and implementation of JavaScript. JavaScript will be used extensively in the website's development as the account system, personalization memory, and contacting are expanded. In doing so, an external service called "EmailJS" had me set up an Email service for the code. In adding the JavaScript, there were issues that came and were resolved, and I am very confident the system will be implemented in a couple days. A signup HTML page was also designed during this period, with a similar layout to the login page.

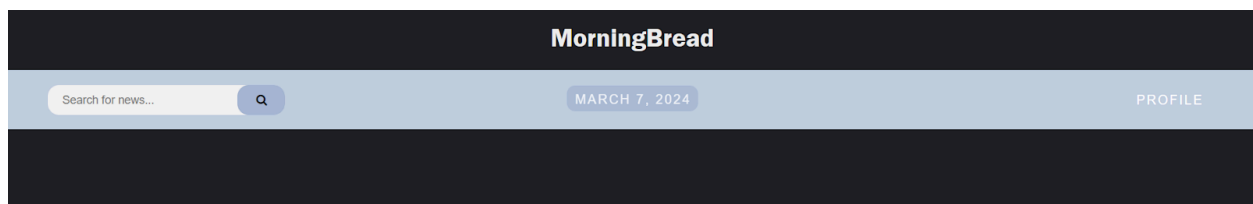
Max-

I worked mostly on the web scraping algorithms last week. This week, I got the SQL database up and running as well as writing the script to upload to the db. I also started researching the NLP side of things and will investigate Latent Semantic Analysis (LSA) as it seems to be the most promising.

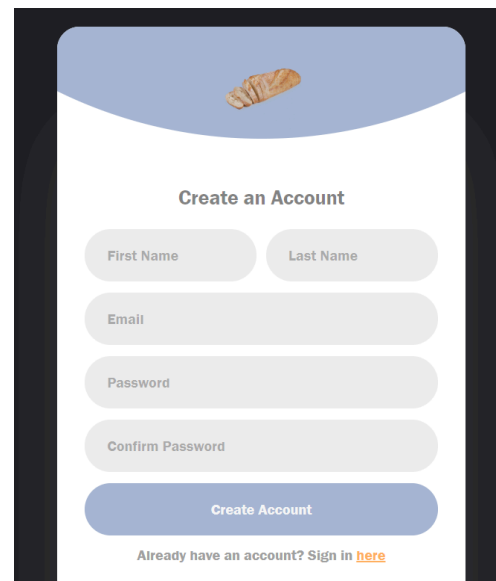
3. **Detailed Progress this Period, separated by Team Member** (Provide detailed information on the progress that you made in the reporting weeks. Include screenshots of code, your game or website, etc. Each team member should have a separate subsection covering their accomplishments. Not including screenshots, this section should be 1-2 pages.)

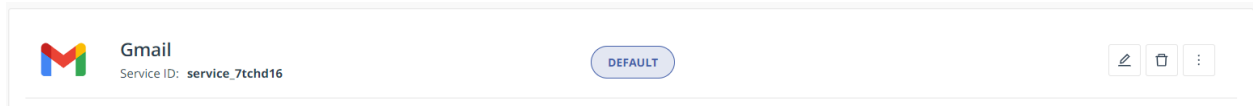
Jake

- I spent a little bit of time in the Codecademy “JavaScript for Web Development” course, but I knew I didn’t want to stick with it for long because getting to work was my main priority
- I needed to fix what I left broken from the previous period, so I spent about a day and a half finally figuring out how to center-align the date in the navigation bar of the main page



- I also changed the color palette of the entire website because one of my team members didn’t like the harsh blue of the previous design
- I also lightened the borders to add a bit of a pop to the elements without it being too obnoxious
- Next, I went to the signup page
- It was very easy to recycle elements from both the login page and the contact page for this, so it did not take very long to develop
- Trying to get the “First Name” and “Last Name” box on the same line took way longer than it should’ve
 - On my first inquiry asking Claude how to fix it, the model figured it out immediately
 - To put into context, I had asked ChatGPT probably a dozen different ways
- I don’t know why the form is going all the way to the bottom of the page instead of cutting off like the login menu, but I will fix that soon
- Secondly, I started the task of trying to implement some JavaScript into the website, particularly when it came to the contact page
- To do this, I had to create an EmailJS account to create a service for the website (This process is pretty similar to Amazon Web Service)





- I added the Service ID and all other necessary parameters into the website through JavaScript and it (kinda) sends emails to my email when you fill out the form now

```
const emailjsConfig = {
  userId: "g0DUuIImNTUn6r7A6",
  serviceId: "service_7tchd16",
  templateId: "template_opam4d4",
};

emailjs.init(emailjsConfig.userId);

function onSubmitForm() {
  const firstName = document.querySelector('input[name="first_name"]').value;
  const lastName = document.querySelector('input[name="last_name"]').value || 'N/A';
  const email = document.querySelector('input[name="email"]').value;
  const message = document.querySelector('textarea[name="message"]').value;

  // Prepare the template parameters
  const templateParams = {
    from_name: `${firstName} ${lastName}`,
    email: email,
    message: message,
  };

  // Send the email using Email.js
  emailjs.send("service_7tchd16", "template_opam4d4", templateParams)
    .then(function(response) {
      console.log("Email sent successfully:", response);
      showThankYouMessage();
    })
    .catch(function(error) {
      console.error("Error sending email:", error);
    });
}
```

- I say kinda because the sender's email is not sending properly, and also will not send the email if the user puts in an email address, so I'm not really sure what's going on there
- But if the user does not implement an email address, I get this email in my inbox:

MoringBread Message From Jake Dorick



Jake <jdorick8@gmail.com>

to me ▾

Hello ,

You got a new message from Jake Dorick:

Test Email

Best wishes,
EmailJS team

Max

- Added another scraping function for marketwatch using their rss feed

```
111 + def scrape_marketwatch_rss():
112 +     #links to rss feeds
113 +     links = [
114 +         "https://feeds.content.dowjones.io/public/rss/mw_topstories",
115 +         "https://feeds.content.dowjones.io/public/rss/mw_realtimeheadlines",
116 +         "https://feeds.content.dowjones.io/public/rss/mw_bulletins",
117 +         "https://feeds.content.dowjones.io/public/rss/mw_marketpulse"
118 +     ]
119 +
120 +     article_list = []
121 +
122 +     #iterate through links, have the same format
123 +     for link in links:
124 +
125 +         page = requests.get(link)
126 +         soup = BeautifulSoup(page.text, 'lxml')
127 +
128 +         articles = soup.find_all('item')
129 +
130 +         for article in articles:
131 +             article_headline = article.find('title').get_text()
132 +             article_link = article.find('link').get_text()
133 +
134 +             try:
135 +                 date_text = article.find('pubDate').get_text()
136 +             except:
137 +                 date_text = article.find('pubdate').get_text()
138 +             article_date = datetime.datetime.strptime(date_text, "%a, %d %b %Y %H:%M:%S GMT") + timedelta(hours = 7)
139 +
140 +             article_list.append(Article(article_headline, article_link, article_date, "marketwatch rss"))
141 +
142 +     return(article_list)
143 +
144 + #method to run all individual methods
145 + def scrape_all(yahoo_scrolldown = 2000):
146 +     article_list = []
147 +     for article in scrape_finviz():
148 +         article_list.append(article)
149 +     for article in scrape_yahoo(yahoo_scrolldown):
150 +         article_list.append(article)
151 +     for article in scrape_marketwatch_rss():
152 +         article_list.append(article)
153 +
154 +     return article_list
```

- I fixed more issues with date retrieval in the finviz function (missing edge cases), also was having a problem where it would grab blog posts along with articles but I resolved the issue
- Created scripts to initialize the db for the first time as well as automatically update it
 - Ideally, I would use myphpadmin or sqlworkbench for db management but I was unable to download



```
16 NLP_algo/sql_master_updater.py
... @@ -0,0 +1,16 @@
1 + import mysql.connector
2 +
3 + articles_db = mysql.connector.connect(
4 +     host = "mysql-2ed0e70f-morningbread.a.aivencloud.com",
5 +     user = "avnadmin",
6 +     password = "AVNS_-1y1cgAxePfkqdPTpji",
7 +     port = 25747,
8 +     database = "morningbread"
9 + )
10 +
11 + mycursor = articles_db.cursor()
12 +
13 + mycursor.execute("SHOW COLUMNS FROM articles")
14 +
15 + for x in mycursor:
16 +     print(x)

12 NLP_algo/sql_updater.py
... @@ -1,7 +1,15 @@
1 #imports
2 from news_scraper import *
3 + import mysql.connector
4
5 + articles_db = mysql.connector.connect(
6 +     host = "mysql-2ed0e70f-morningbread.a.aivencloud.com",
7 +     user = "avnadmin",
8 +     password = "AVNS_-1y1cgAxePfkqdPTpji",
9 +     port = 25747,
10 +     database = "morningbread"
11 + )
12
13 - articles = scrape_all()
14 + mycursor = articles_db.cursor()
15
16 - print(len(articles))
17
18 + insert_sql = "INSERT INTO articles (headline, link, date, source) VALUES"
```

- Finished (for now) updater script below:

```
17 NLP_algo/sql_updater.py
2 from news_scraper import *
3 import mysql.connector
4
5 + #create database connection object
6 articles_db = mysql.connector.connect(
7     host = "mysql-2ed0e70f-morningbread.aivencloud.com",
8     user = "avnadmin",
9
10
11
12
13 mycursor = articles_db.cursor()
14
15 - insert_sql = "INSERT INTO articles (headline, link, date, source) VALUES)"
16 + insert_sql = "INSERT INTO articles (headline, link, date, source) VALUES (%s, %s, %s, %s)"
17 +
18 + articles = scrape_all()
19 + values = []
20 +
21 + for article in articles:
22 +     values.append((article.headline, article.link, article.date, article.source))
23 +
24 + try:
25 +     mycursor.executemany(insert_sql, values)
26 +     articles_db.commit()
27 + except:
28 +     articles_db.rollback()
29 + finally:
30 +     print(mycursor.rowcount, " rows inserted")
```

- I also restructured the file system to be more organized
- Initiated SQL db on remote server

Service	Nodes	Plan	Cloud	Created
 mysql-2ed0e70f MySQL • Running	 Nodes 1	Free-1-5gb 1 CPU / 1 GB RAM / 5 GB storage	DigitalOcean: nyc United States, New York	3 days

- Also read very confusing math papers about machine learning

4. **Difficulties Encountered this Progress Period** (Provide detailed information on the difficulties and issues that you encountered in the reporting weeks. Discuss mitigation strategies for how you got around or plan to get around these issues.)

Jake

Using LLMs to get around hurdles has been extremely helpful, but they just get completely stuck sometimes. It took me right around a day and a half to figure out how to straighten the date to the center of the navigation bar, and I asked just about every LLM out there (ChatGPT, Gemini, Copilot). Claude has been the best LLM I have ever used for code, it solved two big problems I was having with me asking just one query. It does limit the number of questions you can ask per day though, so that stinks. I’m heavily considering paying \$20/month for the premium version.

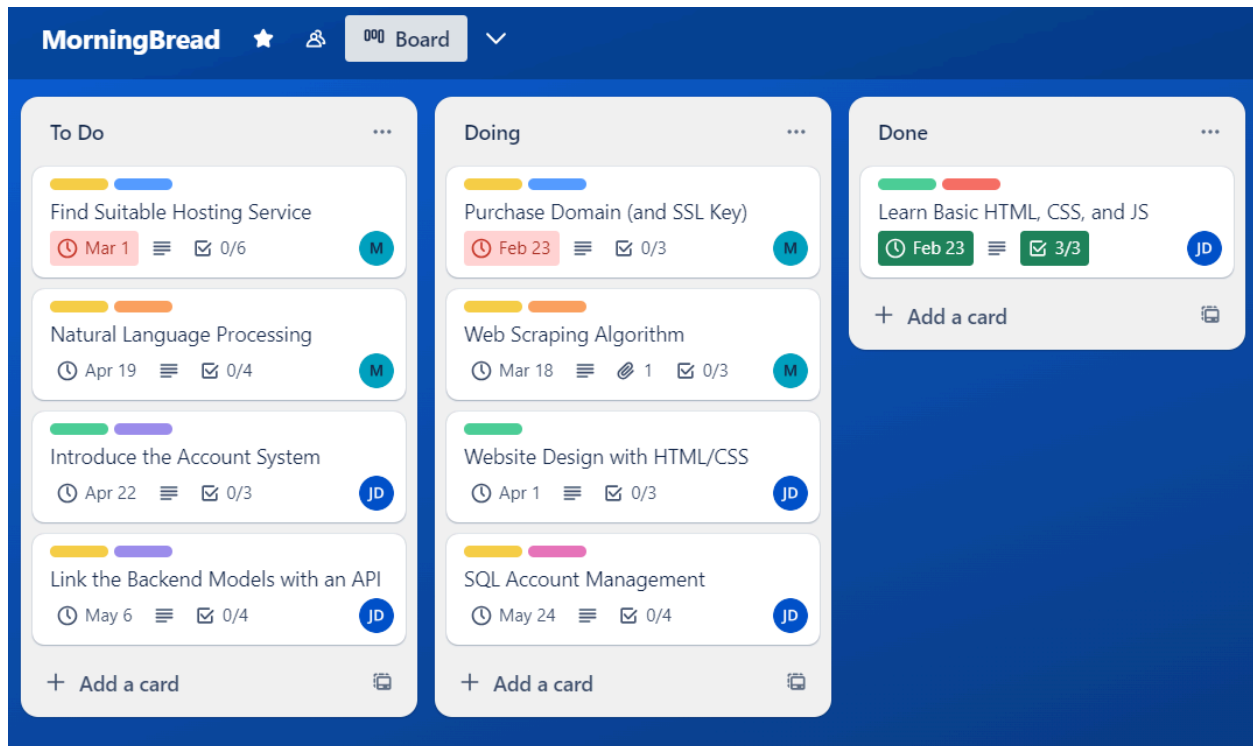
Max

I had no *major* difficulties but there are a few things here and there that I can think of off the top of my head:

- I wasn't able to install software so I had to write a script to manually manage the database
- When I restructured the files, relative imports decided to stop working
- The Yahoo scraping script broke again even though I didn't change anything
- I'm just realizing that ML papers are insanely complicated and use math that I have no hope of understanding

5. **Updated Trello Board and Discussion** (Provide screenshot of and link to updated Trello board. Discuss any changes made to board since last progress report and why.)

<https://trello.com/b/sSuMPdYn/morningbread>



The Trello board needed to be expanded beyond its barebones previous layout. More detailed explanations of website development were added, as well as learned steps of utilizing APIs which will be needed in the future. Due dates were added to those new cards for an appropriate timeline of work.

6. **Tasks to Be Worked on in Next Progress Period** (Discuss the tasks to be worked on in the following two weeks. Discuss who is working on each.)

Jake

I plan to give the frontend development a break for a while. The SQL database of articles needs to be established to give the soon-to-be developed natural language processing model a database to pull from. The knowledge I have from the SQL course I completed for Data Viz plus the refresher during Data Science makes me confident that I will be able to create a functional database as long as the web scraping algorithm works as intended.

Max

Although the scraping methods are not finished, I'm going to move on for now and begin work on the ML side of things. I also need to comment and possibly refactor some code because the code base is growing rapidly and I can see myself getting lost soon in the future.

7. **Additional Information** (Provide any additional information that you want to provide in this section; for example, one of your teammates is going away next week, your Github account is gone, etc. It could be good news as well.)

Jake

I focused a lot on my research project during this time, but I hope it didn't take away the efforts I put into this project. My research gave me a lot of good insight into what we will be doing in the future (especially with APIs and hosting) so I think, even if it did impede some of my work, it will be beneficial in the long run.

Max

I was out for three days so I am slightly behind during this sprint but I think I still got significant work done.
