

Project Name: Materiall

Team members and contact information:

Vincent Lao - vincent.lao@berkeley.edu

Daniel del Carpio - ddelcarpio@berkeley.edu

Bharadwaj Swaminathan - bharadwaj.s@berkeley.edu

Parker Nelson - pnel@berkeley.edu

Samantha Tang - satang1@berkeley.edu

Mentor: Bharat Vijay bharat@materiall.com



iNav

Innovation Navigator Version S1.1
Innovation-Engineering.net

Need	Approach	Benefit	Competition
<p>Who needs it and why?</p> <ul style="list-style-type: none"> • Mainly bigger customers/ businesses looking for specific aspects in a house. • Needed because image postings may or may not have correct tags for features on the house. • Customers may like certain aspects of a house, but not buy it because of others. 	<p>What approach will be used?</p> <ul style="list-style-type: none"> • Image classification to create descriptions of real estate from pictures • Text analysis of given descriptions • Recommendation system for houses 	<p>What is the benefit?</p> <ul style="list-style-type: none"> • Currently, with SIP, people can't go in person to check out houses • Automated system to fill in missing information about homes 	<p>What is the next best alternative if this project was not done?</p> <p>Alternatives: Zillow, ApartmentList</p>

Why is this a good space?	Online project folders, documents:
<p>How will others in the industry react?</p> <p>It will revolutionize the art of homebuying through quantifying the real estate industry through a data science lense.</p>	<p>Google Drive</p> <p>Private Trello for project management (if our mentor is interested)</p>

How will you win?

- Answer:
 - Consistent team meetings
 - Actionable items
 - Clear weekly goals
 - Mutual accountability through active communication

What is Working/Known:

- Answer:
 - Decided on using Octoparse to scrape data from realtor.com
 - Some knowledge of Diffbot and Scrapy in case we need to switch tools

What is Not Working/Known:

- Answer:
 - Need to figure out how to configure the Octoparse bot so that we can easily loop through and scrape all the image urls and text data

Reflection 1:

- Answer:
 - We need to be more consistent on meeting topics
 - So Parker is going to make a meeting agenda for each meeting so that we can keep on track.

Reflection 2:

- Answer:
 - We want to continue creating short-term goals semi-weekly or weekly to make continual progress

Advisor/Manager:

- Answer:

Log Date 4 Oct 2020

10/14 Group Feedback

Detailed Peer Feedback

Score	Team you are Evaluating	Impressive how much they've done already	Seem to be on track to complete this semester	Team needs help	Grand Total
2.4	Materiall	4	6	0	10

Good Parts of Demo

- Project has a very clear knowledge of what they hope to accomplish
- There were a lot of clear visualizations that helped us gain insight into their data.
- They seem to have a very organized way of sharing code on Github.

Areas for Improvement

- I think that they could elaborate more on how each of team members are being productive for the team and the ultimate goal, but this is not as important.
- They are pushing to github right now, and one thing to avoid merge conflict might be to use Google Collab notebook instead
- The code could be a little more descriptive, with text of what the purpose

11/6 Group Feedback

Detailed Peer Feedback

Date	Score	Team you are Evaluating	Impressive how much they've done already	Seem to be on track to complete this semester	Team needs help	Grand Total
10/04	2.4	Materiall	4	6	0	10
11/6	2.2	Materiall	2	9	0	11

Good Parts of Demo

- I really liked how much research they're putting into their work. I think a lot of the time, research is placed second place to more tech-heavy work, but in their case, they are diving deep into theory and pre-existing related projects connected to their project and I value that!
- Have done research on CNNs, which is in addition to the model they have already coded out. So good job doing this research, so next steps would be to implement this model. You have a solid understanding of what steps you need to take next and you're meeting with your mentor today, so great job so far!
- The team has made a lot of progress on creating a pipeline to access their data in an efficient way. It was good that the team was able to quickly identify that their original model was too slow and inefficient so they needed to take another route and have all been researching many different models to find the best one possible

Areas for Improvement

- I think centralizing their visualizations into a powerpoint would be beneficial because it feels like they're bouncing around a lot. Although I also think image classification would be great, it also seems like a big task to implement and it seems like it would be hard to get substantial, consistent results.
- It seems like they are having some issues with storing images because the file sizes are too large. One thing they could look into is various image compression formats such as jpeg2000 or WEBP. Since we are getting to the later stage of the semester, they should really try to make progress on the CNN as soon as possible.
- It seems as though the team has been doing a lot of research for a model/algorithm that would work best for them, but didn't show any example of trial and error. Trying out different models as you research might be more useful to find the best one.

11/11 Group Feedback

Date	Score	Team you are Evaluating	Impressive how much they've done already	Seem to be on track to complete this semester	Team needs help	Grand Total
10/04	2.4	Materiall	4	6	0	10
11/6	2.2	Materiall	2	9	0	11
11/11	2.0	Materiall	0	8	0	8

Good Parts of Demo

- This week they focused on putting images into shapes and running neural networks on the images, which is great progress from last week! They have access to more data now, which is helpful. You guys have made a lot more progress from last week, and have clear steps for what to work on for next week.
- They have a lot of good code down and seem to be on a very good path! It's hard to give feedback on hard code but from what I could see, they have done a lot of hard work and are on a good path to completing for the semester.

Areas for Improvement

- I think adding some visualizations when talking about the data would be more helpful than just talking about it. It's harder to follow without something to look at. I also think that it would be beneficial to give a brief overview of the project in the beginning of the presentation.
- I think it is very smart that they are concentrated on getting the model running first and then planning on fixing+improving+polishing up the nitty gritty later. I hope they get enough time to do so!
- I think adding some visualizations when talking about the data would be more helpful than just talking about it. It's harder to follow without something to look at. I also think that it would be beneficial to give a brief overview of the project in the beginning of the presentation.

How will you win?

- Answer:
 - Frequent communication within the team and with mentors
 - Delegated tasks to organize separate portions of the project

What is Working/Known:

- Answer:
 - Have been utilizing what data we do have to begin preprocessing and initial analysis
 - Will be receiving access to AWS accounts to transfer data to a S3 bucket

What is Not Working/Unknown:

- Answer:
 - Getting blocked by Realtor.com as we continue to scrape data off the web
 - Diffbot and Octoparse have been slightly difficult to use for extracting URLs, text, pics.

Reflection 1:

- Answer:
 - We'll want to start moving forward in our goals - so instead of focusing solely on collecting the data, we will need to start implementing models and continue making progress

Reflection 2:

- Answer:
 - We have weekly team meetings, and will want to continue updating our mentor(s) about our progress through emails and zoom calls

Advisor/Manager:

- Answer:

Log Date : 10/26/2020

End of Project Checklist:

To do list / assigned name:

To do list / assigned name:

