Math​ ​402 ​Final​ ​Project​ ​Peer​ ​Review

Project​ ​Author:​ ​Jeff Hansen, Dylan Skinner, Jason Vasquez, Dallin Stewart

Reviewer:​ Trevor Wai

**Motivation**​ ​**and**​ ​**Overview**​ ​**of**​ ​**the**​ ​**Data:**​ ​**9/10**​ ​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​​(10​ ​very​ ​important​ ​points!)

Is​ ​the​ ​purpose​ ​of​ ​the​ ​project​ ​clear?​ ​What​ ​questions​ ​do​ ​the​ ​project​ ​try​ ​to​ ​answer?​ ​How​ ​could​ ​the objective​ ​be​ ​clarified?

The purpose of their project is clear and they aim to tackle estimating taxi ride durations in NYC based on starting and stopping coordinates. You could possibly elaborate more on what is taxi ride duration, like time of pick up to time of drop off, however I feel this is pretty clear already.

The​ ​author​ ​should​ ​​*briefly*​​ ​review​ ​what​ ​is​ ​already​ ​known​ ​about​ ​the​ ​research​ ​questions​ ​and​ ​what techniques​ ​others​ ​have​ ​used​ ​to​ ​study​ ​these​ ​questions.​ ​How​ ​well​ ​does​ ​the​ ​author​ ​explain​ ​the​ ​scope of​ ​the​ ​project​ ​and​ ​how​ ​it​ ​fits​ ​into​ ​existing​ ​research?​ ​How​ ​could​ ​the​ ​explanation​ ​be​ ​improved? Does the author cite any existing research?

They do a good job of explaining the data set, the breakdown of the dataset, and that it has been used in coding competitions.

I think they could do a better job of explaining the how their approach improves upon existing models as well as citing a few of the existing models.

Does​ ​the​ ​author​ ​give​ ​an​ ​overview​ ​of​ ​the​ ​data?​ ​Do​ ​they​ ​evaluate​ ​the​ ​validity​ ​of​ ​the​ ​data​ ​source? Do​ ​they​ ​defend​ ​why​ ​this​ ​data​ ​set​ ​is​ ​relevant​ ​to​ ​the​ ​project​ ​objective?​ ​What​ ​is​ ​missing​ ​from​ ​the overview​ ​of​ ​the​ ​data?

Yes they give a great overview of the data. I think that they could possibly figure out how Kaggle got the data but I think that they do a fine job of giving an overview and explaining why it is relevant.

**Data**​ ​**Collection**​ ​**and**​ ​**Cleaning:**​ ​**14/15**

The​ ​author​ ​should​ ​explain​ ​how​ ​the​ ​data​ ​was​ ​collected​ ​and​ ​include​ ​any​ ​corresponding​ ​code (possibly​ ​in​ ​an​ ​auxiliary​ ​file).​ ​Is​ ​the​ ​collection​ ​procedure​ ​clear?​ ​Are​ ​there​ ​any​ ​missing​ ​references or​ ​potential​ ​legal​ ​problems​ ​with​ ​the​ ​way​ ​the​ ​data​ ​was​ ​gathered?

Again possibly include how Kaggle gathered the data and mention if there are possible problems of collecting data similar in the future.

The​ ​author​ ​should​ ​explain​ ​how​ ​the​ ​data​ ​was​ ​cleaned​ ​and​ ​include​ ​any​ ​corresponding​ ​code (possibly​ ​in​ ​an​ ​auxiliary​ ​file).​ ​Is​ ​the​ ​cleaning​ ​procedure​ ​clear?​ ​Does​ ​the​ ​data​ ​need​ ​more​ ​cleaning, and​ ​if​ ​so,​ ​how?​ ​If​ ​there​ ​is​ ​little​ ​or​ ​no​ ​data​ ​cleaning,​ ​the​ ​author​ ​should​ ​strongly​ ​defend​ ​why​ ​no cleaning​ ​is​ ​needed.

Yes the way the cleaned the data is very clear.

How​ ​does​ ​the​ ​author​ ​handle​ ​missing,​ ​badly​ ​formatted,​ ​or​ ​incorrect​ ​data?​ ​Does​ ​the​ ​author​ ​justify their​ ​choices​ ​of​ ​what​ ​they​ ​removed,​ ​edited,​ ​reformatted,​ ​or​ ​left​ ​unchanged?

Yes, they include data that they reformatted, creating multiple time features, distance features and weather.

Include the reason why you made these changes. I also think it is really cool you used another Kaggle dataset to make the weather feature.

The​ ​author​ ​should​ ​justify​ ​any​ ​newly​ ​engineered​ ​features.​ ​Are​ ​there​ ​any​ ​additional​ ​features​ ​that should​ ​be​ ​engineered?​ ​If​ ​there​ ​is​ ​little​ ​or​ ​no​ ​feature​ ​engineering,​ ​the​ ​author​ ​should​ ​strongly defend​ ​why​ ​no​ ​feature​ ​engineering​ ​is​ ​needed.

I think that the feature engineering was great especially the pickup period sin/cos. I would also like to see a reason for the feature engineering.

​**Robustness:**​ ​**10/10**

If the author provided code, evaluate​ ​the​ ​readability​ ​of it​ ​throughout​ ​the​ ​project,​ ​especially​ ​code​ ​for​ ​scraping​ ​or cleaning.​ ​ ​Do​ ​they comment​ ​their​ ​code​ ​neatly?​ ​Are​ ​variable​ ​names​ ​reasonable?​ ​Does​ ​every​ ​function​ ​have​ ​an appropriate​ ​docstring?​ ​Can​ ​you​ ​tell​ ​what​ ​the​ ​code​ ​does​ ​just​ ​by​ ​looking​ ​at​ ​it?​ ​How​ ​would​ ​you improve​ ​the​ ​code​ ​to​ ​make​ ​it​ ​more​ ​readable?

Code quality looks great.

How​ ​robust​ ​is​ ​the​ ​cleaning​ ​and​ ​scraping​ ​code?​ ​Could​ ​the​ ​code​ ​be​ ​easily​ ​modified​ ​so​ ​that​ ​it​ ​is usable​ ​for​ ​similar​ ​data​ ​sets,​ ​or​ ​so​ ​it​ ​could​ ​handle​ ​the​ ​same​ ​data​ ​set​ ​but​ ​with​ ​more​ ​data​ ​points? Identify​ ​parts​ ​of​ ​the​ ​code​ ​that​ ​are not​ ​general​ ​enough​ ​and​ ​make​ ​suggestions​ ​for​ ​improving​ ​it.

I think the code for scraping looked good.

**Data**​ ​**Visualization**​ ​**and Basic Analysis:**​ ​**15/15**The​ ​author​ ​should​ ​use​ ​summary​ ​statistics​ ​and​ ​visualizations​ ​to​ ​display​ ​and​ ​describe​ ​the​ ​data​ ​with some​ ​depth.​ ​What​ ​parts​ ​of​ ​the​ ​data​ ​are​ ​still​ ​mysterious​ ​or​ ​unexplained?  
How​ ​well​ ​do​ ​the​ ​visualizations​ ​and​ ​analysis​ ​contribute​ ​to​ ​answering​ ​the​ ​questions​ ​from​ ​the introduction?​ ​Are​ ​there​ ​any​ ​questions​ ​that​ ​are​ ​not​ ​addressed​ ​adequately?

You guys do a great job of explaining the graphs and the graphs look great. I really liked seeing the graph about how weather affected pickup frequency. Personally I didn’t have any unaddressed questions.

Are​ ​the​ ​visualizations​ ​readable?​ ​Do​ ​they​ ​convey​ ​the​ ​information​ ​clearly,​ ​and​ ​in​ ​an​ ​aesthetic way?​ ​Identify​ ​any​ ​visualizations​ ​that​ ​do not​ ​make​ ​sense,​ ​that​ ​should​ ​be​ ​made​ ​with​ ​a​ ​different​ ​kind of​ ​plot,​ ​or​ ​that​ ​can​ ​be​ ​simplified​ ​or​ ​otherwise​ ​improved.

The graphs are all readable. The title font and legends are a little small, but overall the graphs all look really good.

Evaluate​ ​the​ ​validity​ ​of​ ​the​ ​analysis​ ​and​ ​conclusions​ ​about​ ​the​ ​data.​ ​Does​ ​the​ ​author​ ​avoid​ ​bias? Are​ ​any​ ​of​ ​the​ ​conclusions​ ​statistically​ ​unsound?​ ​Point​ ​out​ ​anything​ ​that​ ​you​ ​find​ ​suspicious.

All of the conclusions about the data seem sound. They point out facts that the data show and that would make sense logically thinking about how most people work in downtown and live elsewhere and this corresponds to the times of pickup and drop off as well.

**Learning Algorithms and In-depth Analysis 15/15**

Have the authors used each of the machine learning techniques introduced in the class on the data set? Have they documented the results of their attempts?

Yes they used machine learning techniques introduced in the class and documented each attempt with their model.

For the algorithms that worked well, have the authors provided metrics (e.g., accuracy numbers) to justify that they worked well? Are there any visualizations that should be provided to help document successes? Have the authors commented about why the algorithms gave good results?

You they provide metrics as well, visually and explain it as well. Possibly explain why your algorithms are good, explain more why the evidence suggests the model is less overfit and what that means.

For the algorithms that did not work or were not applicable, have the authors explained why this was the case.

They do a good job of explain the feature space was not large enough and how they could improve.

Have the authors documented which questions the advanced learning algorithms were able to answer? Are the results written up in such a way that this is clear?

Yes they document which questions they were able to answer. They could add explanations why they came to these conclusions.

**Quality**​ ​**of**​ ​**Communication:**​ ​**20/20** Find​ ​and​ ​correct​ ​grammatical​ ​mistakes.

Is​ ​the​ ​presentation​ ​clear​ ​and​ ​easy​ ​to​ ​follow?​ ​What​ ​parts​ ​of​ ​the​ ​project​ ​are​ ​the​ ​most​ ​confusing, and​ ​why?​ ​Make​ ​suggestions​ ​for​ ​improvement.

The project was clear and easy to follow. I think the most confusing part was finding the specific code that you used for different parts because all the code was at the bottom. If possible I would maybe link the functions so that in the top part of the page when you mention them it takes you to where they are in the codebase however I don’t think that is necessary.

Identify​ ​wordy​ ​passages​ ​and​ ​suggest​ ​ways​ ​to​ ​shorten​ ​them.​ ​Where​ ​does​ ​the​ ​author​ ​use​ ​an​ ​entire paragraph​ ​to​ ​say​ ​something​ ​that​ ​could​ ​be​ ​said​ ​in​ ​a​ ​single​ ​sentence?​ ​Are​ ​there​ ​any​ ​$50-words​ ​that could​ ​be​ ​replaced​ ​with​ ​a​ ​50¢-word?​ ​Is​ ​there​ ​any​ ​unnecessary​ ​repetition?

I think they do a great job of keeping conclusions and sentences very short and to the point. I think they could add more explanation behind some of those, but overall I think there isn’t any unnecessary repetition.

**Ethics: 10/ 10**

Did the authors analyze the ethical implications of their research questions, the data gathered, and the analysis performed?  Are there privacy or other implications from the collection or use of the data that the authors did not notice?

They do a good job of analyzing the ethical implications of collecting data. I think that privacy of where people work and where people live is a serious issue of privacy when it comes to collecting data.

Could the results and methods from this project be misused or misunderstood?  Did the authors address this possible misuse?

Yeah they do a good job of addressing possible misuse and consequences.

Is there any other feedback about the ethical implications of the project you wish to give?

They could possibly talk about ethical responsibilities of promoting more effective and efficient taxi traffic could lead to more taxis on the road and the climate effect that could have or the rate of accidents, automotive and pedestrian, increasing or decreasing with their model.

**Impressiveness:**​ ​**10/10**If​ ​you​ ​are​ ​an​ ​employer,​ ​how​ ​much​ ​does​ ​this​ ​project​ ​make​ ​you​ ​want​ ​to​ ​give​ ​the​ ​author​ ​a​ ​job?

I think this was a great creative project. I never would have thought to bring up something like this yet it is something that is extremely practical and demonstrates to me that they are critical thinkers who find common everyday problems and work to solve them.

How​ ​could​ ​it​ ​be​ ​improved​ ​to​ ​make​ ​you​ ​more​ ​eager​ ​to​ ​hire​ ​the​ ​author?​ ​​​Explain.

Overall, I think this project is great and I would love to hire this team.

**Other​ ​comments**:

Great job overall. I think the only ways you could improve are flushing out the reasoning and explanations more.

**Total**​ ​**Score:**​ ​**103/105 (90 for this checkpoint)**