1 Journal Entry for 18 Sep 2023 to 24 Sep 2023

Activities: This week we primarily focused on our strategies for data collection and data management. We again looked at our database schema and what all we wanted to be able to query and manage, and we went from there. We wanted to have data that is reliable and abundant for us to work with so we put a lot of time and thought it what we wanted to do with that. I went through with a lot of different strategies that involved where we could get our data, as though we were set on the SDSS we wanted to see if there was any way we could cross reference any our findings to get more reliable classifications for what we had. I looked at different papers to see their approaches and tried to find catalogs of different types of galaxies, to get an idea of what we could use. At the end I think we figured out that the best course of action is just to use the classifications that the SDSS gives but at lower redshifts. This gives more accurate results and this tactic has been employed by papers that have been published, so I feel like this is a viable strategy that we can undertake going forward. We also edited our paper a little bit more, just putting some finishing touches on it so that we can eventually submit it for publication. I also think that we are in a good position with the initial paper, but still need to do some reworkings with what we have done so far in our data collection.

Challenges: The biggest challenge that we faced this week was regarding our data collection and trying to find viable classifications for our galaxies. It was hard to find a very robust and easy to access catalogue that had good labels for the different types of galaxies besides the SDSS, which is why we settled on using the SDSS. Though the classifications might not be 100 percent accurate which is the main problem we thought about, I think the classification methods in any catalogue we used would be the same, so I think that this is our best option going forward, and we are just going to try to make the existing data we have more accurate by accurately selecting subsets of the data.

Plans: Our plans for this upcoming week involve looking more at data collection, and trying to query data and get an existing working data set. After we do this we will be able to play around with the data and do some initial preprocessing so we can hopefully begin using it for a working model. This is what we are going to have to do going forward, and I am confident that we will be able to figure it all out.

Comments: Overall this week was not too bad and was just focused on data collection and reinforcing what we already have. We are going to really focus down on our topic and understand some of the more nuances in our classification techniques and approaches so that we can continue moving forward.

2 Journal Entry for 11 Sep 2023 to 17 Sep 2023

Activities: This week we revised our review paper based on the comments that were provided to us by our TAs. That was our main focus this week as we wanted to edit and revise our paper to put it in a form that would be viable and ready for publication. We fixed our citations that were not properly put in Latex form, added diagrams to make some points come across more clear as well as revise some of the big problems that were cited in the ideas that we were trying to communicate. Furthermore we began the early stages of getting ready for our next task in trying to consolidate a solid data source where we can solidly extract data from and reach viable conclusions. We created a simple doc that outlines the basic things that we want to address about data collection in respect to our feasibility study. I think we are sticking with extracting flux information from the SDSS but we also want to start accounting for error as well as get a count of how many galaxies we are eventually going to be able to survey. Furthermore, we want to find an accurate classification for our training set in our supervised models, which we aim to look for this week. So in general this week is mostly about consolidating our ideas for data collections and understanding what approaches we wish to take.

Challenges: The biggest challenge that we faced this week was just rewording our paper and deciding how to make it more academic and appropriate for submission. This was not as much a challenge as more of what we focused on for this week and what we spent most of our time this week on. It was not too difficult to do this, but it did require more time and we did run into some difficulties with formatting and Latex but we were able to get everything figured out and working well.

Plans: Our plans for this upcoming week involve looking more at data collection. We want to firmly establish some of the things that we are unsure about so we have more a definite path of what we are doing in the upcoming weeks and have a solid plan that we can end up working on going forward. We want to firmly establish what emission line ratios we are using, what classifications we want to definitively classify into and how we can determine a proper metric to use for our supervised model in determining classifications.

Comments: Overall this week was not too bad and is more of a transition week into what we have already done into what we hope to do going forward. We are going to really focus down on our topic and understand some of the more nuances in our classification techniques and approaches so that we can continue moving forward.

3 Journal Entry for 04 Sep 2023 to 10 Sep 2023

Activities: This week we finished our paper, feasibility study and the presentation that we gave on Thursday. Prior to this week we had been nearly done with our paper but we put the finishing touches on it and put it in accordance with the research template document we were given. We were able to wrap that up and write it in an academic style that we hope will help us get published in an undergraduate journal. Furthermore we worked on our feasibility study more, to get an idea of what were able to work with going forward with our research. We developed some basic models but realized that we need to spend a lot more time thinking about how we can collect data and properly manipulate it so that we can work with it. In the end being able to parse the data and determining how to preprocess it will probably be the most important that we will need to take into consideration. Furthermore we developed our presentation and were able to present it. We spent time preparing the presentation so we could effectively communicate what we were trying to do and get the best feedback we could. I think that we will take the things Dr. Gebhardt told us very seriously and try to address some of the problems we are currently facing in terms of data collection and metrics to check for supervised classification.

Challenges: The biggest challenge that we faced again was our feasibility study and also managing time. There was a lot that we had to do this week and just being able to do it all while managing our other time commitments was difficult but we were able to get it done. We had to put the final touches on our paper, presentation, and feasibility study all in the same week but in the end it worked, and we have more to go off of as we continue our research. The feasibility study for this week posed some difficulties such as trying to figure out the SVM model and get TensorFlow working. However by just messing around a little bit with it, it was not too bad and we got it figured out.

Plans: Our plans for this upcoming week involve looking more at data collection and maybe even trying a different database that Dr.Gebhardt suggested. We need a better metric to check our supervised model against because the metric provided by the SDSS is not 100 percent accurate and that is something we need to address. From there we want to more solidly define what emission line galaxies we want to classify. Right now we know we want to classify, Seyfert 1, 2, LINERs, and HII regions but we want to make sure we are sticking to this and not thinking about any more or any less.

Comments: Overall this week was a lot of work but I think we got a lot done which will help us going forward in the upcoming weeks and develop a more solid research focus. I think we do need to put more time into how we are going to collect our data and what methods we might use to better document the type of galaxies we are training with in our training set, and

also what we really want to accomplish in our use of a K-Means Model.

4 Journal Entry for 28 Aug 2023 to 03 Sep 2023

Activities: This week we continued to write our first paper that gave an overview of our field recounting what we are doing, the history of the field, current challenges, and future direction. I have finished writing my sections and we are going to compile it and put the whole thing together so we will definitely be able to turn in the paper in time. Furthermore, we started working on our presentation for the upcoming week, essentially creating a slideshow that covered a lot of the big points that we talk about in our paper as well as our feasibility study. The presentation is essentially finished so we are ready to be able to present it in the upcoming week. What took the most time this week was expanding on the feasibility study code we had worked on the week before. We now have 3 very basic models from data we queried from the SDSS, a SVM Model, a K-Means Model and a Neural Network. All of the models seem to be giving positive results and show signs of things that we want to see, which makes it seem that our reserach problem is feasible to move off from. Though our supervised models are based of of BPT predictions, which is one thing that we want to address in the future, potentially finding another metric to check against to classify for a supervised model.

Challenges: The biggest challenge faced this week was also with code again. It was difficult getting some of the models to work but though it took time we were able to come with some results. At first I had struggled with getting the K-Means model to work, however my group mates were able to help me and we got some visualization that could prove to be handy for investigate purposes and also a way of forming labels for a supervised model. The paper writing was not too difficult, I had pretty much been finished with my part and now we are just trying to link everything together and make it all cohesive.

Plans: For next week, I hope to finalize our paper and presentation. We are almost done with both so we just need to put finishing touches on everything but we should be good to go. The work that we are doing right now is just laying the foundation for everything and as we progress throughout the semester we will now have a basis to actually build a proper model with hopefully better supervised labels that we can use. All in all though preparing for the presentation will be the biggest thing we will have for this upcoming week.

Comments: Overall this week was a solid amount of work with writing the paper and working on the feasibility study. I am looking forward to being able to develop our Machine Learning models and think that we got a really good foundation and idea of the field through writing the initial paper and messing around with data collection and training models. I think that we are

in a good place to get a good model and a good research paper out of what we are doing.

5 Journal Entry for 21 Aug 2023 to 27 Aug 2023

Activities: This week I continued writing the first paper to build up a foundation of our research. I found more sources, read them and inserted them into Zotero and was able to develop a more solid understanding of what I was learning. I made a lot of progress this in week of terms of our first paper and I think I have a very solid understanding of the history of our field as that was the task I was assigned to writing for our paper. I think that there is a lot of existing literature on the subject but with the rise of modern computing there is a lot more to do, so I think our topic is a good balance of both and will yield a good research paper. Furthermore, we started developing a simple prototype for our model to gauge its feasibility. We were able to query data from the SDSS and only took a small sample for the initial purpose to see some level of success, and currently we are working on it to get some results. We started out with a K-Means model and we are going to try to develop that but we might switch over to a SVM model as it might more accurately reflect how galaxy emission-line classification works. We will continue working on that up to the next meeting, and continue developing the paper even more but I think we have a pretty good hang of it going into it now.

Challenges: This week the biggest challenges I faced was with my code. I had some trouble getting the K-Means model to work and processing and filtering data. I did get past the initial preprocessing state with data, and am now trying to effectively fit it to a K-Means model and visualize it. However, the K-Means model might identity alternate patterns then what we are trying to discover so it might be more advantageous to switch to a SVM model which is what we should gauge. The paper writing is not posing too much of a challenge, it is just time consuming, but I do find it interesting to go through reserach papers and talk about what I am getting from them.

Plans: For next week, I continue to keep refining our paper and try to get a working model of our prototype. We just want a model that classifies galaxies to some level of success, so that we can keep expanding on that model throughout the semester. It would provide a good basis for working off of and would show us that our project is feasible to some extent. We will also keep writing our paper and attempt to getting a working rough draft by this week so we can assemble a solid final copy by the due date.

Comments: Overall this week was a solid amount of work with writing the paper and developing a simple prototype for our model. I think that it will get a little less stressful as we continue throughout the semester, and establish more of an idea on how we are continuing on

our research and what results we are getting out of everything. Also, it may take some more time and effort to get some code working but we should be able to figure it out this week.