AAI-530 Final Team Project Details

Instructions:

In the Final Team Project for this course, you will work with a team of 1-2 of your classmates (assigned by your instructor) to design and implement a machine learning IoT application. You will work closely with your teammates to find an existing IoT dataset, create a complete (theoretical) IoT system design for that dataset/application, use machine learning methods to provide insights from the raw sensor data, and present those insights on a Tableau Public dashboard.

When you complete this project, you will have applied all of the skills and knowledge that you have gained in this course, and you will have a good idea of what is involved in designing and implementing an IoT system in the real world. Your submitted report, GitHub repository, and Tableau Public dashboard will be good portfolio items when applying for machine learning roles in the future.

Final projects and datasets from prior courses cannot be used. You and your team are required to use GitHub as a code hosting platform to manage version control and collaboration during this project. Your code and your repository must be well-documented, including a README file that describes all files in the repository and comments throughout all code files.

Datasets:

You and your team will need to identify an IoT dataset to use in your project. This dataset can be from any type of IoT system, including smart home data, health data, wearable devices, smart city data, etc. The only requirements for your IoT dataset are:

- It must come from real (not simulated) IoT sensor/device(s).
- It cannot be the Household Energy Consumption dataset or the Air Quality or Continuous Glucose Monitoring datasets as these are used heavily in the course assignments and lab sessions.
- If you use a dataset from the textbook, you cannot use the same machine learning method that it is used to demonstrate in the textbook.
- It must include at least one time series variable.

Some possible sources for datasets are listed below:

Hands-On Artificial Intelligence for IoT Textbook
Github UC Irvine Machine Learning Repository
25 Datasets for Deep Learning in IoT

IoT System Design:

Use the knowledge that you have gained from this course to design a complete (theoretical) IoT system for the IoT application dataset that you have chosen. Similar to Assignment 1, your system design should include **a diagram of all system components with accompanying documentation** of what each component is and how it contributed to the system design. *This design does not need to be a report of the actual system used to collect and analyze your chosen dataset*, but you should use what you have learned to design a realistic system for your chosen IoT application.

Items your System Design must address include:

- **Sensors** - Indicate the location, type, and any limitations of your sensor(s).
- **Edge processing** - Does your system require any computation on the edge? How does this affect the requirements for your edge device(s)?
- **Networking** - How does your device connect to the network? What messaging protocol does it use to communicate?
- **Data storage and processing** - What tools and storage systems are used to handle the incoming device data? How will your system manage scalability? How and where are your machine learning insights produced?

Your diagram and documentation should be included in your submitted report.

Data Processing and Insights:

Your submitted code files should include a brief exploratory analysis (similar to Assignment 2) with any data cleaning/processing that you performed on your IoT dataset.

Your submitted code files should also include the code used to perform at least two different machine learning tasks on your IoT dataset. You may use any methods covered in this or any of your other courses, with the requirements that your methods must include:

- **At least one using deep learning** - This means that you must build and train a deep learning method from scratch using Tensorflow or another Python machine learning library (e.g. Pytorch). You may not use a pre-built model architecture or a pre-trained model to fulfill this requirement.
- **At least one time series prediction** - Similar to Assignments 3 and 4, this task must create a prediction for the future based on current/past data.

So, your two models could be:

1. a deep learning classifier (like a CNN object detector) and a traditional ML time series predictor (like a linear regression model for predictive maintenance);
2. two deep learning time series predictors (like an LSTM to predict soil moisture and an LSTM to predict daily temperature);
3. or a deep learning time series predictor (like a transformer model that predicts hourly glucose levels) and a traditional ML classifier (like an SVM for identifying heart murmurs)

Your two models must predict different variables, so training two different models to make the same prediction will not satisfy t

Include a brief description of the two methods that you chose in your submitted report.

IoT Dashboard:

Similar to Assignment 5, your submitted Tableau Public Dashboard must include:

- **At least one status visualization** - This visualization should tell the dashboard user something about the "current" status of their IoT device (e.g. number of devices online, current glucose level, etc.)
- **At least one summary visualization** - This visualization should tell the dashboard user something about the historical data from their IoT device (e.g. average device downtime over the last week, number of hypo/hyperglycemic readings over the last week)
- **At least one visualization for each of your two machine learning insights** - This visualization should communicate the insights created by your machine learning methods.

In designing your dashboard and visualizations, remember the best practices covered in Module 5. In particular, your visualizations should:

- Use pre-attentive attributes to draw attention to the most important information.
- Share a consistent color scheme.
- Be organized on the dashboard in a logical order.
- Have clean and clear titles and labels.

Project Timeline:

- **Module 2 (by the end of Week 2):** The course instructor will group students into teams of two to three members. Each team should select and introduce a dataset and IoT application/system of their choice. Canvas, USD Email, or Slack can be used to find your team members.
- **Module 4 (by the end of week 4):** Your team will submit a status report of the project. One team representative will need to submit a .pdf which briefly describes the machine learning methods your team has chosen.
- **Module 7 (by the end of Week 7):** One team representative should submit the three deliverables described below for the course project in the final week.

** It is critical to note that no extensions will be given for any of the final project due dates for any reason, and final projects submitted after the final due date will not be graded.

Project Deliverables and Submission Format:

- **Report**: Submit a **PDF** document containing a final report with an explanation of your IoT system design, the machine learning methods you chose, and the design choices that you made on your dashboard. The report should be in APA 7 style (sample APA professional paper) and should be 10-15 pages (double-spaced), including text, IoT system diagram, and selected tables/graphs.

- **Code**: Submit a **PDF** (for notebooks) or .py file(s) with the code for your data processing and two machine learning methods. Also, include a link to the GitHub repository with all of the final code files merged. Your code should be well-commented such that your instructor can tell what each piece is doing without a lot of effort.

- **Dashboard**: Submit a **PDF** and a **link** to your dashboard on Tableau *Public* with at least four visualizations (described above) relevant to your chosen IoT application.

Final **PDF** documents and your team's GitHub and Tableau Public links must be submitted through this assignment link. **Only <u>one member</u> of your team will need to submit these deliverables.**

Plagiarism, or passing another person's work off as one's own either by directly copying or even paraphrasing it without proper citation, is a serious offense and can result in sanctions including grade reductions, course failures, and even expulsion from the university. For more information, please see the USD Code of Honor.

Scoring Rubric

Criteria	**Achievement Level Meets or Exceeds Expectations**	**Achievement Level Approaches Expectations**	**Achievement Level Below Expectations**	**Achievement Level Inadequate Attempt**	**Achievement Level Non-Performanc e**
IoT System Design 20%	All components of the IoT system are present on the diagram and clearly described in the accompanying documentation.	Most components of the IoT system are present on the diagram and/or described in the accompanying documentation.	Some components of the IoT system are present on the diagram and/or described in the accompanying documentation.	Few components of the IoT system are present on the diagram and/or described in the accompanying documentation.	IoT system diagram and/or documentation is missing.
Data Processing 10%	Exploratory data analysis and cleaning script is complete and well-commented.	Exploratory data analysis and cleaning script is mostly complete and/or includes some comments.	Exploratory data analysis and cleaning script is somewhat complete and/or includes very few comments.	Exploratory data analysis and cleaning script is incomplete and/or includes no comments.	Exploratory data analysis and cleaning script is missing.
Machine Learning Insights 40%	Machine learning methods include at least one deep learning model and one	Machine learning methods include at least one deep	Machine learning methods include at least one deep	Machine learning methods do not include at least one deep	Machine learning methods are missing.

	time series prediction. The code is clean and well-commented.	learning model and one time series prediction. The code is difficult to follow or missing comments.	learning model or one time series prediction. The code may be difficult to follow or missing comments. OR The code is only partially functional.	learning model or time series prediction. The code may be difficult to follow or missing comments. OR The code is non-functional.	
Tableau Dashboard 20%	All four required visualizations are present and organized in an effective way to communicate the major data insights.	All four required visualizations are present, but are not organized with data communication in mind.	Two or three visualizations are present, and are organized in an effective way to communicate major data insights.	Two or three visualizations are present, but are not organized with data communication in mind.	Dashboard is missing or only has one visualization.
Writing Mechanics 10%	Report demonstrates quality graduate level writing skills. Student posts are free of spelling and grammatical errors. Resources and references are appropriately cited in APA format (where applicable).	Report has some spelling and grammatical errors and needs some revision to meet expectations. Resources may not be appropriately cited in APA format (where applicable).	Report has many spelling and grammatical errors and needs major revision to meet expectations. Resources are not appropriately cited in APA format (where applicable).	Submission does not meet graduate level standards.	Non-performance