MEMORANDUM

Date: \_\_ 23FEB19\_\_\_

From:  LT Rouben Azad

Section(s): Computer Science (368)

To:  Program Officer, Computer Science: LCDR Eric Regnier

Via:  (1) Thesis Advisor: Dr. John V. Monaco

(2) Co-Advisor or Second Reader: Dr. Geoffrey Xie

(3) Academic Associate, CS Department: Dr. Alan Shaffer

(4) Chair, CS Department: Dr. Peter J. Denning

Subj: THESIS PROPOSAL

Encl: (1) Computer Science Thesis Proposal

(2) Institutional Review Board (IRB) Student Research Checklist

1. Tentative Title of Proposed Thesis: The Impact of Machine Learning on AIS Data.

2. General Area of Proposed Thesis Research: This thesis will show how a machine learning algorithm can be used to detect abnormal activities captured by AIS data.

3. Enclosure (1) is the Thesis Proposal with a milestone plan of dates and events for research and thesis completion.

4. I expect that my thesis will be unclassified.If classified, I have read Chapter VIII of NPSINST 5510.2F and the NPS Research Admin web page concerning classified theses (https://my.nps.edu/web/thesisprocessing/classified-fouo-papers)

5. I reviewed the IRB webpage concerning the use of humans in research (https://my.nps.edu/web/research/irb-home). This research does NOTinvolve human subject research.

6. I anticipate the following travel or other extraordinary requirements: None.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student Signature

1. Forwarded, recommending approval: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_

Thesis Advisor Date

2. Forwarded, recommending approval: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_

Co-Advisor or 2nd Reader (circle) Date

3. Forwarded, recommending approval: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_

Academic Associate, CS Dept Date

4. Forwarded, recommending approval: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_

Chair, CS Department Date

5. Approved, and retained:  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_

Program Officer, CS Department Date

COMPUTER SCIENCE THESIS PROPOSAL

A. General Information

1. Name: LT Rouben Azad

2. Email: rlazad@nps.edu

3. Curriculum: Computer Science (368)

4. Thesis Advisor: Dr. John V. Monaco

5. Co-Advisor: Dr. Geoffrey Xie

6. Academic Associate, CS Department: Dr. Alan Shaffer

7. Chair, CS Department: Dr. Peter J. Denning

8. Date of Graduation: March 2020

B. Area of Research

This research will demonstrate how a machine learning algorithm can effectively identify a wide variety of anomalies in vessel dimensions, identification, ship to ship networking, and movement behaviors in an Automated Identification System (AIS) database. By analyzing this data through machine learning, we can show a whole new field of possibility for maritime security by supplying them with additional tools to enable better security procedures out at sea and ports. Machine learning is now being used across all fields of technology and the United States Department of Defense can benefit greatly by this implementation. Sea-going vessels are notorious for using spoofing technologies to alter AIS data. This thesis will involve development of an algorithm that can spot these anomalies and categorize vessels as seen fit.

The title of this thesis will be: The Impact of Machine Learning on AIS Data.

C. Research Questions

Hypothesis: Machine learning can be used to detect and categorize AIS data anomalies.

1. Can machine learning be used to detect anomalies in AIS data better than current technologies?
2. What is the impact of machine learning on maritime security?
3. What are the advantages and disadvantages of this algorithm?
4. Can this algorithm be used to survey other systems similar to AIS?
5. Can this be used to identifies vessel, their operators, country of origin or class by their navigation behaviors?
6. Is it possible to detect covert communication between ships with this algorithm?

D. Discussion

An Automated Identification System (AIS) [2] is the primary source of identification for surface vessels on the high seas. Almost all ships are equipped with them, and they are used frequently by other vessels to aid in collision avoidance. The information that AIS provides, such as, unique identification, course, position, speed, tonnage, and ship dimensions, is especially useful in and around ports. Port controls use this data to direct traffic and maintain channel control. Maritime authorize and asses this data, and physically verify the vessels for security, or otherwise.

Although this data is relied on heavily by all factors listed above, and more, it is easily “spoofed” -- meaning, it can be altered before transmission. Some vessels use spoofing to hide their identity or ship loadout. Security risks are on the rise due to frequent use of spoofing data from sea-going vessels, especially in this technological age where data altering is common. [1]

Artificial intelligence (AI) and machine learning have been a new facet of discussion. This fairly new technique can have a powerful impact on information sharing across vessels because it has the ability to predict and prevent security risks at sea.

Our main steps in building a machine learning algorithm are; preparing the data, choosing our model, training, evaluating, tuning and prediction. This will be discussed further in the methodology section. [6]

AIS data will be used to train a model to make predictions and decisions without explicit programming. The goal is to find outliers in this data by the machine’s use of pattern recognition, classification, and regression algorithms, which will point us to our anomalies.

This machine learning algorithm will parse the AIS data to detect and categorize anomalies which will help maritime authorities gain deeper knowledge on risky vessels.

Most vessels out at sea don’t have fraud detection that can prevent against spoofing. Automated anomaly detection could potentially aid in collision avoidance. By gathering information on high-risk vessels, it also enables watch standers to make better decisions at sea.

E. Scope of the Thesis

This thesis will show how a machine learning algorithm can be used to detect abnormal activities extracted from this AIS data. This algorithm will be able to categorize vessels by their maneuvers, behaviors, and network connections by using pattern recognition, observing anomalies, and finding outliers.

F. Methodology

This thesis will involve working with programming languages such as python, with additional modeling interfaces for analyzing and assessing data. The tools that will be utilized for this work are IPython for its interactive shell commands that support data visualities and graphical user interface (GUI) toolkit; and Python’s Pandas and Matplotlib, which are both great tools for data analysis and modeling. Matplotlib will be primarily used for plotting libraries in its 2D plotting interface.

The first step in the process is data preparation, by loading our data into a suitable place. Then begin parsing the data using an established algorithm with a few modifications to support the changing data structures. After completion of the parsing and organization, a model will be constructed that suits the data. Afterword’s we will train our model with the given data, and incrementally improve its ability to make better predictions. Once the training is complete evaluation beings to test the model’s accuracy. From the evaluations we can decide to further improve our training by tuning our parameters. And lastly the prediction step, we use our model to predict the correct answers. [6]

This machine-learning algorithm will be constructed using the previously stated methods. There will be multiple levels of AI to be established in this structure, beginning with data analyzation using 2D models with Matplotlib; machine-learning acquisition and testing; data categorization; and ultimately, anomaly detection of real-world historic AIS data.

The second step will be analyzing large amounts of data over a long period of time using the first step. Afterwards, the algorithm will be put through validation to narrow down errors and complete the program.

Finally, a result of the stress tests will be documented and improved upon. In turn, this program will be used to analyze data in real time to provide a quick response to maritime security.

G. Chapter Outline

1. Introduction
2. Background
   1. Current uses of machine learning and AI in this field
   2. Rules of the road on the high sea and etiquette
   3. AIS background and technologies
   4. Benefits of this research
3. Machine Learning Techniques
   1. Techniques and uses for the algorithm
   2. Parsing algorithm
   3. Data modeling
4. Prototype Systems
   1. Parsing AIS data
   2. Database construction
   3. System architecture and development
5. Exploratory analysis
   1. Analysis findings
6. System Testing and Analysis of Results
   1. Task analysis
   2. Debugging and error analysis
7. Discussion
   1. Impacts on maritime security
8. Conclusion and Summary
   1. Areas of Future Research
9. Appendices
10. Bibliography

H. Schedule

|  |  |
| --- | --- |
| ***Thesis Stage*** | ***Completion Target Date*** |
| Literature review | MAY 19 |
| Draft thesis, cover through Chapter 2, and initial list of references <to advisor(s) to check background knowledge and problem statement, and to approve proposed plan to conduct research> | MAY 19 |
| Draft thesis checked by thesis processor for format | JUN 19 |
| Construct research design | JUL 19 |
| Conduct research and any associated travel | AUG 19 |
| System testing and stability | AUG 19 |
| Analyze data | SEP 19 |
| Draft thesis to advisors | NOV 19 |
| Final thesis submission for signatures | DEC 19 |

I. Benefits of Study

If our hypothesis is correct, this work will help maritime authorities gain deeper knowledge on high-risk vessels and help eliminate misinformation and illegal activities at sea. This will also equip watch standers with an improved situational awareness aided by the use of this algorithm, which will enhance the information received by AIS for better decision making and judgment control.

J. Anticipated Travel/Funding Requirements

None, as of yet.

K. Preliminary Bibliography

1. Kyle Wilhoit and Marco Balduzzi, “Vulnerabilities Discovered in Global Vessel Tracking Systems.” *TrendLabs Security Intelligence Blog*, 8 Aug. 2014, blog.trendmicro.com/trendlabs-security-intelligence/vulnerabilities-discovered-in-global-vessel-tracking-systems/.
2. “What Is the Automatic Identification System (AIS)?” *MarineTraffic Help*, Marine Traffic, help.marinetraffic.com/hc/en-us/articles/204581828-What-is-the-Automatic-Identification-System-AIS-.
3. Natalija Jolić, Mato Šimić, and Sanja Brnadić-Zoranić. “Strategy of Implementing Automated Identification System -ITS in Port Systems.” *Promet (Zagreb)* 15.5 (2003): 333–337. Web.
4. Marco Balduzzi, Alessandro Pasta, and Kyle Wilhoit, “A Security Evaluation of AIS Automated Identiﬁcation System.” *A Security Evaluation of AIS Automated Identiﬁcation System*, 2 Jan. 2015, www.iseclab.org/people/embyte/papers/ais\_acsac14.pdf.
5. Lynch, Robert. “EMC ANALYSIS OF UNIVERSAL AUTOMATIC IDENTIFICATION AND PUBLIC CORRESPONDENCE SYSTEMS IN THE MARITIME VHF BAND .” *Transition*, Feb. 2004, transition.fcc.gov/omd/dataquality/peer-reviews/ais/source-1.pdf.
6. “IEC TECHNICAL COMMITTEE 80: MARITIME NAVIGATION AND RADIOCOMMUNICATION EQUIPMENT AND SYSTEMS.” INTERNATIONAL ELECTROTECHNICAL COMMISSION IEC. *Making Electrotechnology Work for You*, 2011, www.iec.ch/about/brochures/pdf/technology/maritime\_nav\_rc\_2.pdf.
7. Yufeng, and Yufeng G. “The 7 Steps of Machine Learning.” *Towards Data Science*, 31 Aug. 2017, towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e.
8. Kaluza, Pablo, et al. “The Complex Network of Global Cargo Ship Movements.” *Royal Society Publishing,* 19 Jan. 2010, royalsocietypublishing.org/doi/pdf/10.1098/rsif.2009.0495.
9. Cull, Kimberly M. “Classifying Vessels Operating in the South China Sea by Origin with the Automatic Identification System.” *Calhoun*, Naval Postgraduate School, Mar. 2018, calhoun.nps.edu/bitstream/handle/10945/58289/18Mar\_Cull\_Kimberly.pdf.
10. Liraz, Shay Paz. “SHIPS’ TRAJECTORIES PREDICTION USING RECURRENT NEURAL NETWORKS BASED ON AIS DATA.” *Calhoun*, Naval Postgraduate School, Sept. 2018, calhoun.nps.edu/bitstream/handle/10945/60431/18Sep\_Liraz\_Shay\_Paz.pdf.
11. Hintze, Jhon R. “An Analysis of Vessel Waypoint Behavior through Data Clustering.” *Calhoun,* Naval Postgraduate School, Sept. 2017, calhoun.nps.edu/bitstream/handle/10945/56135/17Sep\_Hintze\_John.pdf.

L. Data Management Plan

This data will be archived in the NPS GitLab repository.

*M. Preliminary Abstract*

Automated identification system data is utilized commonly in the open sea and in ports. It is also easily “spoofed” -- meaning, it can be altered before transmission. Some vessels use spoofing to hide their identity or ship loadout. Security risks are on the rise due to frequent use of spoofing data from sea-going vessels, especially in this technological age where data altering is common.

This research will focus on building and training a model through machine learning that will be used to make predictions and decision support.  The goal is to construct a ship-to-ship network through the AIS database which is comprised of; ship coordinates, dimensions, identification and characteristics. The algorithm will be used to detect changes and abnormal activities over a given period of time.

Most vessels out at sea don’t have fraud detection that can prevent against spoofing.  This algorithm we develop will learn and detect changes via the data it receives, which could potentially aid in collision avoidance.  By gathering information on high-risk vessels, it enables watch standers to make better decisions at sea.