Adversarial Machine Learning (December 2018)

Sarp Aykent, Jordan Cox, and Blake Schilleci

***Abstract*—** **Now that we have broken our algorithm and found our flaws, we are not prepared to fortify our algorithm and defend against any attack. With this idea in mind, a way that our best algorithm was tested was by allowing 3rd parties to manipulate/mask an article from our dataset to attempt a breach and expose our algorithm’s inaccuracies. In every sense of the phrase, this is Adversarial Machine Learning at its core. Being able to test our algorithm against unknown mask is the best way to indicate how strong it really is.**

# INTRODUCTION

From our last challenge, we attempted to break our machine learning algorithm to expose its flaws and how inconsistent it really is. Handling any attack on our algorithm would be very difficult if we did not know the issues that our method held. A way to make anything stronger is to force it to fail. Anything under pressure will in turn, come out stronger. In this project, we are allowing our machine learning algorithm to be exposed even more than how we exposed it. When dealing with 3rd parties who are attacking your algorithm, the thought processes and methods that they use to break our code is going to be different than anything that we might think of. This creates a big challenge and allows our algorithm to really see how strong it is. Along with this, we were able to attack our adversaries with masked articles to attempt to break their system.

# Methodology

Our methodology for attacking was simple. We wanted to take the articles from other teams and our own CASIS/SEC authors [1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21

22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 38 39 40 41 42

43 44 45] and change them just enough to throw off the system, but we also wanted our attacks to be hidden. To do this, we used an online software for translation through Google and Yandex. This is simple at surface level but the programs that we use to mask the articles do heavy lifting in order to alter the samples.

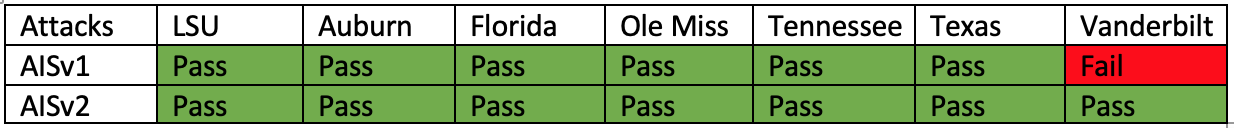
Our methodology for defense was also simple. We didn’t want to over complicate it by adding unnecessary features that could possibly make our accuracy worse after our articles were masked, so we focused on analyzing the decision function when our system was identifying for authorship.

# Experiment

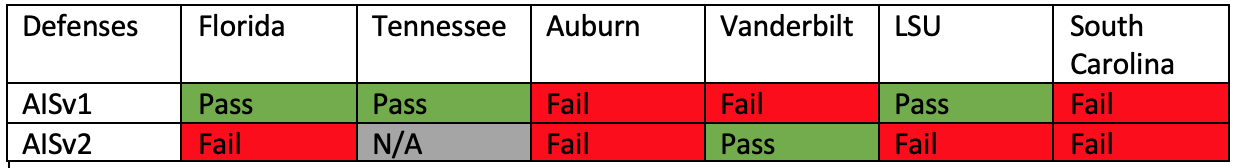
When attacking, we would take the writing samples provided us and put them through an online translator. When we translated, we chose to translate to Chinese then Russian and back to English. We chose these languages for specific reasons. We chose Chinese because they use a very large amount of characters within their language, and their language has very simple grammar without tense when constructing sentences. From there, we converted to Russian, which also features unique characters from Chinese and English. Russian also has a very complex grammatical structure juxtaposed to Chinese. This allowed us to make a robust attack.

For the actual experiment with defending against attacks, our approach was simple yet effective. Creating a threshold for the decision function allowed our system to determine what was an attack and what wasn’t. Making the threshold 0 allowed for anything other than an attack to return a positive number, and anything under zero(negative) to be flagged as an attack and skipped. After the attack is found, the rest of the articles are then ran through the decision function and correct predictions are made. By ignoring the obvious attacks, our system is still able to predict the authors at a high rate and also not be fooled by blatant attacks.

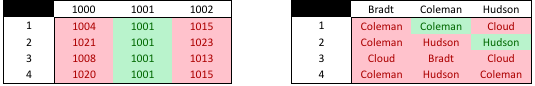
# Results



Above is all attacks done on to our system.



Above is all defenses done against our attacks.



Above is how well our system defended against our own attacks.

# Breakdown of Work

**Sarp Aykent:** Attacked and defended. Co-wrote paper.

**Jordan Cox:** Attacked and defended. Co-wrote paper.

**Blake Schilleci:** Attacked and defended. Co-wrote paper.

1. References

[1] A. Press. “No. 18 Mississippi State tops SFA 63-6 in Moorhead's debut.” *ESPN,* 1 Sept. 2018. [Online]. www.espn.com.

[2] A. Press. “Fitzgerald leads no. 18 miss st to 31-10 rout of K-State.” *ESPN,* 8 Sept. 2018. [Online]. www.espn.com.

[3] A. Press. “No. 18 Mississippi State breezes by SF Austin.” *SEC Sports,* 8 Sept. 2018. [Online]. www.secsports.com.

[4] A. Press. “No. 18 Mississippi State stifles Kansas State 31-10.” *SEC Sports,* 8 Sept. 2018. [Online]. www.secsports.com.

[5] H. Cloud. “Defensive dog collars, slam dunks, are all part of a culture change.” *The Reflector,*12 Sept. 2018. [Online]. www.reflector-online.com.

[6] H. Cloud. “Bulldog offense lifts off, defense smothers in first win for Moorhead.” *The Reflector,*2 Sept. 2018. [Online]. www.reflector-online.com.

[7] J. Coleman. “Bulldogs becoming disruptive on defense.” *Bulldogs Becoming Disruptive on Defense | Starkville Daily News*, 2 Sept. 2018. [Online]. Starkvilledailynews.com.

[8] T. Horka. “Running dack Kylin Hill leads Mississippi State to 31-10 victory over Kansas State.” *The Clarion Ledger*, The Clarion-Ledger, 8 Sept. 2018. [Online]. www.clarionledger.com.

[9] T. Horka. “Thompson, Hill lead Mississippi State to 63-6 stomping of Stephen F. Austin.” *The Clarion Ledger*, The Clarion-Ledger, 2 Sept. 2018. [Online]. www.clarionledger.com.

[10] B. Hudson. “MSU's defense shows versatility in dominating Kansas State.” *The Commercial Dispatch*, 10 Sept. 2018. [Online]. www.cdispatch.com.

[11] B Hudson. “Columbus' Hill has rousing start to season.” *The Commercial Dispatch*, 2 Sept. 2018. [Online]. www.cdispatch.com.

[12] D. Skretta. “Thrill from Hill: sophomore running back produces in MSU victory.” *Bulldogs Becoming Disruptive on Defense | Starkville Daily News*, 8 Sept. 2018. [Online]. starkvilledailynews.com.

[13] D. Brandt. “Fitzgerald, No. 16 Mississippi State Rout La.-Layayette.” *ESPN*, ESPN Internet Ventures, 16 Sept. 2018, www.espn.com/college-football/recap?gameId=401012278.

[14] D. Brandt. “Snell Leads Kentucky Past No. 14 Mississippi State 28-7.” *ESPN*, ESPN Internet Ventures, 23 Sept. 2018, www.espn.com/college-football/recap?gameId=401012286.

[15] D. Brandt. “Mullen, Florida Beat No. 23 Mississippi State 13-6.” *ESPN*, ESPN Internet Ventures, 30 Sept. 2018, www.espn.com/college-football/recap?gameId=401012298.

[16] D. Brandt. “Fitzgerald Leads Miss St to 23-9 Win over No. 8 Auburn.” *ESPN*, ESPN Internet Ventures, 7 Oct. 2018, www.espn.com/college-football/recap?gameId=401012304.

[17] J. Coleman. “A-Train Gets Back on Track for No. 14 MSU.” *A-Train Gets Back on Track for No. 14 MSU | Starkville Daily News*, 16 Sept. 2018, starkvilledailynews.com/node/124495.

[18] G. Graves. “Bad Night: Wildcats Put Damper on Plans of Bulldogs.” *Bad Night: Wildcats Put Damper on Plans of Bulldogs | Starkville Daily News*, 23 Sept. 2018, starkvilledailynews.com/node/124552.

[19] J. Coleman. “Searching for Answers: Bulldogs Left Scratching Their Heads as Former Coach Leaves with a Win.” *Searching for Answers: Bulldogs Left Scratching Their Heads as Former Coach Leaves with a Win | Starkville Daily News*, 30 Sept. 2018, starkvilledailynews.com/node/124604.

[20] J. Coleman. “Sweet Relief: MSU Finds End Zone, Victory over Auburn.” *Sweet Relief: MSU Finds End Zone, Victory over Auburn | Starkville Daily News*, 7 Oct. 2018, starkvilledailynews.com/node/124650.

[21] B. Hudson. “Fitzgerald Looks Dominant as No. 16 Bulldogs Slice and Dice Ragin' Cajuns.” *The Commercial Dispatch*, www.cdispatch.com/sports/article.asp?aid=68524.

[22] B. Hudson. “Kentucky Upsets No. 14 MSU.” *The Commercial Dispatch*, www.cdispatch.com/sports/article.asp?aid=68661.

[23] B. Hudson. “MSU's Offense Struggles in Loss to Florida.” *The Commercial Dispatch*,

www.cdispatch.com/sports/article.asp?aid=68856.

[24] B. Hudson. “Running Games Propels MSU Past No. 8 Auburn.” *The Commercial Dispatch*,

www.cdispatch.com/sports/article.asp?aid=69007.

[25] H. Cloud. “Wilcox Honored, MSU Rolls to Victory in Third Game of Season.” *The Reflector*, 19

Sept. 2018, www.reflector-online.com/sports/article\_c849b1bc-b977-11e8-9f39-3faf1e8637ad.html.

[26] H. Cloud. “Bulldogs Knock down No. 9 Tigers.” *The Reflector*, 8 Oct. 2018, www.reflector-

online.com/sports/article\_bf4c2cc4-c9ef-11e8-8bb1-8b906830601d.html.

[27] H. Cloud. “Moving on from Kentucky, MSU Focused on First Home SEC Game.” *The Reflector*, 27

Sept. 2018, www.reflector-online.com/sports/article\_a15ade6a-c25e-11e8-bb32-ab7b74da8d2a.html.

[28]H. Cloud. “Mullen's Return Proves Unhappy for Bulldogs.” *The Reflector*, 30 Sept. 2018, www.reflector-online.com/sports/article\_36296bcc-c469-11e8-af8e-eb779ea01df7.html.

[29] H. Cloud. “Making a statement in win over Aggies,” The Reflector, 30-Oct-2018. [Online]. Available: http://www.reflector-online.com/sports/article\_6767fa6e-daf8-11e8-bf0b-7b9c303aff93.html. [Accessed: 04-Nov-2018].

[30] H. Cloud, “MSU falls silently as offense dies in Death Valley,” The Reflector, 22-Oct-2018. [Online]. Available: http://www.reflector-online.com/sports/article\_af69ae1c-d4e4-11e8-ac39-5b3a9e9c1c4d.html. [Accessed: 04-Nov-2018].

[31] B. Hudson. “LSU delivers Death Valley demolition,” The Dispatch, 22-Oct-2018. [Online]. Available: http://www.cdispatch.com/sports/article.asp?aid=69313. [Accessed: 04-Nov-2018].

[32] B. Hudson. “Fitzgerald (4 TDs) leads MSU past No. 16 Texas A&M,” The Dispatch, 30-Oct-2018. [Online]. Available: http://www.cdispatch.com/sports/article.asp?aid=69500. [Accessed: 04-Nov-2018].

[33] J. Coleman. “Going about business: MSU's Fitzgerald silences critics, leads MSU to big home win,” Starkville Daily News, 22-Oct-2018. [Online]. Available: http://www.starkvilledailynews.com/node/124802. [Accessed: 04-Nov-2018].

[34] J. Coleman. “Bright spot: Despite loss, MSU defense continues to play well,” Starkville Daily News, 22-Oct-2018. [Online]. Available: http://www.starkvilledailynews.com/node/124756. [Accessed: 04-Nov-2018].

[35] B. Martel. “Defense leads No. 5 LSU past No. 22 MSU, 19-3,” ESPN, 20-Oct-2018. [Online]. Available: http://www.espn.com/college-football/recap?gameId=401012315. [Accessed: 04-Nov-2018].

[36] D. Brandt. “Mississippi State upsets No. 16 Texas A&M, 28-13,” ESPN, 27-Oct-2018. [Online]. Available: http://www.espn.com/college-football/recap?gameId=401012321. [Accessed: 04-Nov-2018].

[37]A. Narayanan, H. Paskov, N. Z. Gong, J. Bethencourt, E. Stefanov, E. C. R. Shin, and D. Song, *On the Feasibility of Internet-Scale Author Identification - IEEE Conference Publication*. [Online]. Available: https://ieeexplore.ieee.org/document/6234420. [Accessed: 04-Nov-2018].

[38] D. Brandt. “Fitzgerald throws 4 TDs, Mississippi St beats La Tech 45-3,” ESPN, 04-Nov-2018. [Online]. Available: http://www.espn.com/college-football/recap?gameId=401012329. [Accessed: 14-Nov-2018].

[39] D. Brandt. “Defense powers No. 1 Alabama past No. 18 MSU, 24-0,” ESPN, 11-Nov-2018. [Online]. Available: http://www.espn.com/college-football/recap?gameId=401012331. [Accessed: 14-Nov-2018].

[40] J. Coleman. “Win No. 6: MSU reaches bowl eligibility with blowout victory,” Sidewalk discord continues | Starkville Daily News, 04-Nov-2018. [Online]. Available: http://www.starkvilledailynews.com/node/124854. [Accessed: 14-Nov-2018].

[41] J. Coleman. “Situation goes wrong: Top-ranked Tide too much for Bulldogs,” Sidewalk discord continues | Starkville Daily News, 10-Nov-2018. [Online]. Available: http://www.starkvilledailynews.com/node/124915. [Accessed: 14-Nov-2018].

[42] B. Hudson. “No. 21 Mississippi State routs Louisiana Tech,” The Commercial Dispatch. [Online]. Available: http://www.cdispatch.com/sports/article.asp?aid=69650. [Accessed: 14-Nov-2018].

[43] B. Hudson. “No. 18 MSU far from perfect against No. 1 Alabama,” The Commercial Dispatch. [Online]. Available: http://www.cdispatch.com/sports/article.asp?aid=69787. [Accessed: 14-Nov-2018].

[44] H. Cloud. “Home sweet homecoming victory,” The Reflector, 05-Nov-2018. [Online]. Available: http://www.reflector-online.com/sports/article\_05020174-dfeb-11e8-a911-afa8bce3dc65.html. [Accessed: 14-Nov-2018].

[45] H. Cloud. “Starkville native steps up in game against No. 1 Alabama,” The Reflector, 12-Nov-2018. [Online]. Available: http://www.reflector-online.com/sports/article\_22099fda-e620-11e8-9fb5-bbc165b8a603.html. [Accessed: 14-Nov-2018].