IRB #: IRB-FY23-24-929
Title: Detecting Government Fraud Using Semi-Supervised Machine Learning
Creation Date: 3-9-2024
End Date:
Status: Approved
Principal Investigator: William Sapp
Review Board: NU IRB
Sponsor:

Study History

Submission Type	Initial	Review Type	Exempt	Decision	No Human Subjects Research

Key Study Contacts

Member	Lawrence Fulton	Role	Co-Principal Investigator	Contact	lfulton@nu.edu
Member	William Sapp	Role	Principal Investigator	Contact	w.sapp6678@o365.ncu.edu
Member	William Sapp	Role	Primary Contact	Contact	w.sapp6678@o365.ncu.edu

Initial Submission

Before you get started review and complete all steps outlined on our [website](#).

*required

Acknowledge you have completed all of the following steps prior to starting this application:

All boxes must be checked to move on

- ✓ [CITI Training](#)

- ✓ Review [institutional policies](#) about research

- ✓ Consider the [feasibility](#) of your study

- ✓ Prepare [supplemental documents](#)

*Helpful Resources:

1. Attend [Group Writing Sessions](#) for additional help with supplemental documents and/or IRB Application

2. Review the [How to Complete an IRB Application](#) video

3. Review the [How to Submit and Certify your IRB Application video](#)

3. Review [IRB Process Infographic](#)

4. Review [Submission Statuses Guide](#)

Who will be the Primary Contact for this project?

STUDENTS: Select yourself.

FACULTY: Select yourself as primary contact unless you have a Research Coordinator who will be making submissions on your behalf.

STAFF: Select yourself as primary contact unless you have a Research Coordinator who will be making submissions on your behalf.
Name: William Sapp
Organization: NCU-ST-Technology-phd-ds
Address:
Phone:
Email: w.sapp6678@o365.ncu.edu

What is the full title of the research study?

Detecting Government Fraud Using Semi-Supervised Learning

Study Personnel

Study personnel refers to anyone who will be *analyzing* study data. Study personnel does **NOT** include transcription services or statisticians.

Who is the Principal Investigator (PI)?

The PI is almost always the person responsible for collecting and analyzing data. There can be only one PI listed on the study who is ultimately responsible for the conduct and oversight of the study including education of study staff, study management, and the protection of study participants.

Name: William Sapp
Organization: NCU-ST-Technology-phd-ds
Address:
Phone:
Email: w.sapp6678@o365.ncu.edu

*required

Upload CITI certification

The current requirement for CITI training is the **Social & Behavioral Educational (SBE)** course (former NCU researchers)
or the **Researchers - Social and Behavioral Focus** course (researchers that have always been affiliated with NU or any study personnel that are not affiliated with NU).

Please ONLY upload the required CITI training certificate, we do not need to review any additional certificates you've completed.

Click [here](#) for CITI User Guide
[citiCompletionCertificate_12631456_58410416.pdf](#)

*required

At NU:

I am Faculty

✓ I am a student

*required

Student researchers must have a Faculty Sponsor/Chair certify their submission before it comes to the IRB.

Please insert your Faculty Sponsor/Chair's name by clicking the find people button below.

For a quick tutorial video on how to enter the Faculty Chair's name using the Find People button, [click here](#)

Name: Lawrence Fulton
Organization: NCU-School of Technology & Engineering
Address:
Phone:
Email: lfulton@nu.edu

*required

It is the PI's responsibility to ensure any person with access to data has completed required human subjects research training.

Confirm your Faculty Chair has completed the required CITI training by clicking "Yes".

✓ Yes

I am staff

*required

Do you have a Co-Investigator (Co-I)?

Co-PIs/Co-Is are key personnel who have responsibilities similar to that of a PI on research projects. While the PI has ultimate responsibility for the conduct of a research

project, the Co-PI/Co-I is also obligated to ensure the project is conducted in compliance with applicable laws and regulations and institutional policy governing the conduct of research.

Yes

✓ No

Do you have any additional study personnel?

Study personnel can include research assistants, or other individuals who may have <u>access to your data</u> (this does NOT include professional transcription services or statisticians).

TIP: *For PI's working with <u>**unaffiliated Co-Investigators**</u>, please select "yes" here and list the name(s).*

Yes

✓ No

Is this project funded?

Yes

✓ No

Multi-Site IRB Research

- If you are using a research site that has an IRB click here for more information about the requirements

- If your site is not registered with the U.S. Department of Health and Human Services (HHS), Office for Human Research Protections, the answer to this question should be No.

- Many K-12 school districts have an IRB/HRPP office, but they are not a registered IRB with the HHS. If your study will take place in a K-12 setting, the response to this question is No.

*required

Are you conducting your research at a site with an IRB or HRPP (Human Research Protection Program)?

Yes

✓ No

I need NU IRB approval first

International Research

This study involves international research (i.e., research done outside the US)

International research can include collecting data from people outside the US, or using data sets that did not originate in the US. For more information about conducting international research, click here.

Yes

✓ No

Projected Start Date

VERY IMPORTANT: In advance of initiating *any* involvement of human subjects in this study (i.e., contact, recruitment, and enrollment), the IRB must review and accept your response to all of the comments included in the preliminary review.

You may not advertise your study, recruit participants, or collect data until approval for the use of human subjects is obtained.
If you've already started research activities (recruitment, consent, data collection, etc.), please email irb@nu.edu with guidance on how to move forward.

I will wait until I have received IRB approval to initiate *any* involvement of human subjects in this

✓ study (i.e., contact, recruitment, and enrollment)

I have already started contacting participants and/or have collected data from my participants.

*required
What is/are your research question(s)?

TIP: The research question is the overall question your study is trying to address. It reflects the phenomenon you are studying, and what you are attempting to learn about it.

For more information and examples, click on the question mark icon on the right.
RQ1:
What are the key features or variables associated with fraudulent loan applications within the Paycheck Protection Program?
RQ2:
What novel combination of existing supervised and unsupervised machine learning techniques can effectively identify fraudulent activity within the Paycheck Protection Program?

*required
In 2-3 sentences describe your research design.

For more information and examples, click on the question mark icon on the right.
This study will use a quantitative experimental design to identify key features and methodologies to identify Paycheck Protection Program (PPP) loan fraud using supervised and unsupervised machine learning. This quantitative experimental study will use publicly available PPP loan data publicly available from the US Small Business Administration (SBA) and a novel combination of existing unsupervised and supervised machine learning algorithms to detect fraudulent PPP loan applications.

*required
Involvement of Human Subjects

Check all that apply

✓ Archival data

Archival data may have been collected for research purposes or non-research purposes, and from publicly available sources (such as websites), or non-public sources (such as hospital or school records). Researchers often analyze data that were previously collected by someone else. For additional information on using archival, secondary or existing data sets, click [here](#).

*required
Select which type of data set below

Public data set

✓ • Public data sets can be accessed by anyone and do not require permission from the owner(s) of the data

*required
Provide a link to the dataset(s) below

If you are conducting a meta-analysis, instead of providing a link, describe body of published literature to be analyzed
https://data.sba.gov/dataset/ppp-foia

Private data set

Both public and private data sets

Interaction and/or intervention with human subjects

Intervention

Common forms of data collection that involve intervention with human subjects include:

- Physical procedures (e.g.: fatigue stress tests)
- Manipulations of the subject's environment that are performed for research purposes (e.g.: response induction techniques)

Interaction

Common forms of data collection that involve interaction with human subjects include:

- Observing (in places where an individual could reasonably expect not to be)
- Surveying (by instrument or questionnaire)
- Interviewing (systematically across all participants)