**[](http://www.gmu.edu/)**

**Department of Health Administration and Policy**

**College of Health and Human Services**

|  |
| --- |
| **HAP 789, HEALTH INFORMATICS**  **CAPSTONE PROJECT PROPOSAL** |
| **Student Name: Rabia Endris** |
| **Organization: Not affiliated (SFARI Autism BrainNet dataset)** |
| **Title of Project: Predicting Early Autism Diagnosis Using Socioeconomic and Behavioral Data: A Machine Learning Approach** |
| ***Problem Statement/ Statement of Need and Scope of Project***  *Describe the specific organizational issue, problem or opportunity that this capstone project is expected to address. Please address the following items:*   * What is the problem site? (i.e. organization, department, section, etc.) * What problems or issues are you going to address? * Did you have a hypothesis that you plan to prove or disprove? * Did you have an objective that you plan to accomplish? * Are there any limitations regarding the project scope? |
| **Problem Statement/Statement of Need and Scope of Project**   * **Problem Site:** Remote research conducted from home, using the SFARI Autism BrainNet dataset (pending approval) for autism screening and diagnosis research. * **Problems or Issues Addressed:** Delays in autism spectrum disorder (ASD) diagnosis due to socioeconomic disparities (e.g., household income, parental education, healthcare access) and variability in behavioral data (e.g., developmental screenings, parent-reported behaviors). * **Hypothesis:** A machine learning model using socioeconomic and behavioral data from the SFARI Autism BrainNet dataset can predict ASD in children under five with at least 80% accuracy. * **Objective:** Develop and validate a predictive model to identify children at high risk for ASD, supporting prioritized diagnostic evaluations in research settings. * **Limitations:** Limited to retrospective Autism BrainNet data (pending SFARI approval), which may have demographic or geographic biases. No real-time data collection or intervention implementation. Public datasets (e.g., National Survey of Children’s Health, NSCH) will be used if SFARI approval is delayed. |
| ***Expected Impact/Value Added of Project***  *Why is this project important to your capstone organization and what value is anticipated from your work? Please address the following points:*   * *What opportunities or problems are driving the need for this project at this time?* * *Specifically, how will your project address these opportunities or problems?* * *What value do you expect to deliver to your capstone organization as a result of completing this project?* |
| * **Opportunities or Problems Driving the Need:**   + Rising ASD prevalence (1 in 36 children, CDC 2023) and socioeconomic barriers delay diagnosis, limiting early interventions critical before age five.   + Need for predictive tools using accessible research data to enhance ASD diagnostic efficiency. * **How the Project Addresses These Opportunities or Problems:**   + Uses Autism BrainNet data to identify key ASD predictors, enabling prioritization of high-risk cases.   + Supports research to reduce diagnostic delays and disparities. * **Expected Value Delivered:**   + A validated predictive model to contribute to autism research by improving early ASD diagnosis.   + An interactive dashboard to visualize risk factors for researchers and public health professionals.   + A comprehensive report with recommendations for equitable screening practices using research data. |
| ***Research Methods/Methodology to be employed***  *Provide a detailed description of how you will conduct your project including the collection and analysis of data, and the development of conclusions and recommendations that are supported by reliable and valid information. Please address the following items:*   * What research methods will you use? * How will you collect the necessary information? * What analytical methods will you use? * What statistical methods will you use? * How will you design your study to assure validity and reliability of the findings? * How will you design your study to optimize the opportunity for organizational “buy in”? |
| **Research Methods/Methodology to be Employed**   1. **Research Initiation:** Review the SFARI Autism BrainNet dataset (pending approval) to identify variables (socioeconomic: income, education, insurance; behavioral: developmental scores, parent reports). Study prior ASD research and predictive models for methodological guidance. 2. **Data Collection:** Access de-identified Autism BrainNet data via SFARI’s secure protocols on a personal secure laptop compliant with data privacy standards. Use NSCH data as a contingency if SFARI approval is delayed. 3. **Analytical Methods:** Preprocess data (handle missing values, encode categorical variables, normalize numerical data). Apply recursive feature elimination for feature selection. Train supervised machine learning models (logistic regression, random forest, gradient boosting) with a 70-30 train-test split and 5-fold cross-validation. 4. **Statistical Methods:** Evaluate model performance using accuracy, precision, recall, F1-score, and AUC-ROC. Perform feature importance analysis with SHAP values. Use t-tests or chi-square tests to assess variable associations. 5. **Validity and Reliability:** Ensure validity using Autism BrainNet’s DSM-5 diagnostic criteria. Enhance reliability through data cleaning, cross-validation, and comparison with prior ASD studies. Attempt external validation with NSCH data if feasible. 6. **Stakeholder Engagement:** Engage autism research communities by sharing findings through virtual presentations. Design a user-friendly dashboard to facilitate model adoption in research workflows. |
| ***Final Project Deliverable(s)***  *Identify what the final deliverable will be. (i.e. project plan, departmental assessment, facility assessment, business plan, feasibility study, request for proposal, system design recommendations, physician satisfaction study, manpower plan, policy manual, etc.)* |
| **Final Project Deliverable(s)**   * **Predictive Model:** Validated machine learning model for early ASD diagnosis with performance metrics and SHAP-based feature importance analysis. * **Interactive Dashboard:** User-friendly dashboard (e.g., Tableau or Python-based) to visualize ASD risk factors and model predictions. * **Comprehensive Report:** Detailed report on methodology, findings, and recommendations for integrating the model into autism research workflows. * **Data Insights and Recommendations:** Summary of key findings with actionable recommendations for prioritizing diagnostic resources and improving screening equity. |
| **Project Timetable (including components, assignment of responsibility)** |
| |  |  | | --- | --- | | Task | Target Completion Date | | Receive approval for project scope and methodology | July 7, 2025 | | Collect and review relevant background information | July 10, 2025 | | Gain organization approval for data collection method and process | July 14, 2025 | | Complete data collection process | July 21, 2025 | | Complete analysis of data | August 4, 2025 | | Complete first draft of report | August 11, 2025 | | Review conclusions and recommendations with Preceptor | August 14, 2025 | | Submit Final Report | August 18, 2025 | | Prepare Final Presentation | August 21, 2025 | |  |  | |  |  | |
| **Statement that Preceptor has reviewed and approved this Capstone Project Description** |
| This Capstone Project Proposal is pending review and approval by Dr. Eman Elashkar, Faculty Advisor, George Mason University. |