Assignment - 4

Yoginder Singh & Rohan Pradeep

ADMN5016-F23-101: Applied Artificial Intelligence and Machine learning.

Project: - Stroke Prediction

Professor: Sujoy Paul

Dec: 15th, 2023

**Table of contents**

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Particular** | **Page No.** |
| 1. | Problem the Application Solves & Market Size: | 3 |
| 1.1 | Above the Project | 3 |
| 1.2 | **Application Purpose** | 3 |
| 1.3 | **Criticality of Stroke Prediction** | 3 |
| 1.4 | **Market Size** | 4 |
| 2. | Performance Evaluation | 5 |
| 2.1 | Matrix used | 5 |
| 2.2 | Model comparison | 5 |
| 3. | Monetary Value and Risks After Performance | 6 |
| 3.1 | Savings Estimate | 6 |
| 3.2 | Loss Estimate | 7 |
| 3.3 | **Employee Reduction and Efficiency** | 7 |
| 4. | **Other Benefits & Risks** | 8 |
| 5. | **Conclusion** | 8 |
| 6. | **References** | 9 |

**1 Problem the Application Solves & Market Size**:

**1.1 About the Project**

According to the World Health Organization (WHO) stroke is the 2nd leading cause of death globally, responsible for approximately 11% of total deaths. This dataset is used to predict whether a patient is likely to get stroke based on the input parameters like gender, age, various diseases, and smoking status. Each row in the data provides relavant information about the patient.

* 1. **Application Purpose**:
  + **Predictive Analysis**: The application uses machine learning to analyze various factors like gender, age, hypertension, heart disease, marital status, work type, residence type, average glucose level, BMI, and smoking status. This comprehensive analysis aims to predict the likelihood of a stroke, a major health concern globally.
  + **Early Detection**: By identifying high-risk individuals, the application can facilitate early intervention strategies, potentially reducing the severity or even preventing strokes.
  + **Personalized Healthcare**: The application caters to personalized health care by considering individual health profiles, which can lead to more effective and targeted healthcare strategies.
  1. **Criticality of Stroke Prediction**:
  + **Global Health Concern:** Stroke is a leading cause of death and long-term disability worldwide. Early prediction can significantly impact patient outcomes by enabling preventative measures.
  + **Burden on Healthcare Systems:** Strokes require extensive medical care, rehabilitation, and support services, which strain healthcare resources. Predictive tools can help in resource allocation and management.

**1.4 Market Size:**

1. **Global Reach**:
   * **Healthcare Settings**: Hospitals, clinics, and other healthcare facilities globally can use this application for patient assessment and monitoring.
   * **Telemedicine and Mobile Health**: With the rise of telemedicine, such applications can be integrated into remote health monitoring systems, widening the reach.
2. **Demographic Targeting**:
   * **High-Risk Populations**: Regions with higher incidences of stroke or populations with prevalent risk factors (like hypertension, obesity) represent key market segments.
   * **Aging Populations**: Countries with aging populations may find particular value in such predictive tools, given the increased stroke risk with age.
3. **Economic Impact**:
   * **Cost Savings**: Early stroke prediction can lead to cost savings in healthcare by reducing the need for extensive treatment, rehabilitation, and chronic care.
   * **Insurance and Health Policy**: Insurance companies and health policymakers can leverage such tools for risk assessment and to design better health plans.
4. **Integration with Existing Health Systems**:
   * **Data-Driven Healthcare**: Integrating with electronic health records (EHR) systems, the application can enhance data-driven decision-making in healthcare.
   * **Collaboration Opportunities**: Partnerships with health tech companies, hospitals, and research institutions can expand the application's scope and effectiveness.

In summary, the application has the potential to significantly impact global healthcare by providing a tool for early stroke prediction. Its market extends across healthcare settings in various regions, especially those with high stroke prevalence or at-risk populations. The economic impact in terms of cost savings and improved health outcomes further underscores its market potential.

1. **Performance Evaluation**

**2.1 Metrics Used:**

1. **Recall**: Measures the ability of the model to correctly identify all actual positive cases (in this case, patients at risk of stroke). A high recall indicates fewer false negatives (missing out on identifying true stroke risks).
2. **Precision**: Indicates the proportion of positive identifications that were actually correct. High precision means the model has fewer false positives (wrongly identifying a stroke risk).
3. **F1 Score**: Harmonic mean of precision and recall. An F1 score balances both metrics and is particularly useful when the class distribution is imbalanced, as is often the case in medical diagnostics.
4. **Accuracy**: Overall, how often the model predicts correctly, both true positives and true negatives.

**2.2 Model Performance Comparison:**

1. **Logistic Regression**:
   * **High Recall**: This suggests that the Logistic Regression model is quite effective in identifying patients who are at risk of stroke. However, high recall alone might not always be desirable, especially if it comes at the cost of precision.
   * **Low Precision and F1 Score**: This indicates a significant number of false positives, meaning the model often predicts a stroke risk when there isn’t one. In a medical context, this could lead to unnecessary anxiety and medical interventions.
   * **Implication**: While the model is good at detecting true positives, its tendency for false alarms might limit its practical applicability without further refinement.
2. **RandomForestClassifier**:
   * **Better Balanced Performance**: This model demonstrates higher accuracy and F1 score compared to Logistic Regression. It means that it not only correctly identifies a higher proportion of true positive and negative cases but also maintains a better balance between precision and recall.
   * **Higher Accuracy (0.88 vs. 0.77)**: This suggests that overall, the RandomForest model makes more correct predictions (both true positives and true negatives) than the Logistic Regression model.
   * **Higher F1 Score (0.24 vs. 0.12)**: This improved score indicates a better balance between recall and precision, making it a more reliable model in scenarios where both false positives and false negatives carry significant consequences.
   * **Implication**: RandomForestClassifier is potentially more suitable for practical use in a medical setting, given its balanced performance in correctly identifying stroke risks and avoiding false alarms.
3. **Monetary Value and Risks After Performance:**

**3.1 Savings Estimate:**

1. **Prevention of Stroke-Related Costs**:
   * **Direct Healthcare Costs**: Hospitalization, medication, rehabilitation, and long-term care costs can be significant for stroke patients. By preventing strokes, the application can save these direct healthcare expenses.
   * **Indirect Costs**: This includes lost productivity due to disability or death. Stroke is a leading cause of serious long-term disability, which impacts workforce participation and economic productivity.
2. **Estimating Savings**:
   * **Example Scenario**: If the application prevents 50 potential strokes, consider the average cost of stroke treatment per patient. This can range from acute care hospitalization costs to long-term rehabilitation and care expenses. The total savings would be this cost multiplied by 50.
   * **Broader Economic Impact**: Beyond direct cost savings, there's a positive economic impact from maintaining a healthier, more productive workforce.

**3.2 Loss Estimate:**

1. **Cost of False Positives**:
   * **Unnecessary Medical Interventions**: False positives could lead to unnecessary medical tests, treatments, or even psychological stress for patients incorrectly identified as at high risk.
   * **Quantifying Costs**: Estimating this cost involves considering the average cost of unnecessary medical examinations and treatments per false positive case.
2. **Balancing Risk and Benefit**:
   * **Risk-Benefit Analysis**: While false positives carry costs, these need to be weighed against the potential costs and impacts of false negatives (failing to identify a true stroke risk).

**3.3 Employee Reduction and Efficiency:**

1. **Reduction in Manual Analysis**:
   * **Labor Costs**: Automated stroke prediction can potentially reduce the workload of healthcare professionals involved in risk assessment, leading to savings in labor costs.
   * **Efficiency Gains**: Automation can also streamline the process, leading to faster decision-making and potentially earlier interventions.
2. **Estimating Staffing Savings**:
   * **Cost Per Analysis**: Consider the average time and cost for a healthcare professional to conduct a risk assessment. Savings would be this cost multiplied by the number of assessments the application can automate.
   * **Long-Term Implications**: While there might be initial costs in implementing and maintaining the application, the long-term savings from reduced staffing needs and increased efficiency could be substantial.

**4. Other Benefits & Risks**

**Risks:**

1. **Misdiagnosis**: Inaccurate predictions leading to unnecessary treatments or missed stroke cases.
2. **Over-Reliance**: Excessive dependence on the model could undermine clinical judgment.
3. **Data Privacy**: Handling sensitive patient data poses significant confidentiality and security risks.

**Benefits:**

1. **Early Detection**: Facilitates prompt intervention, potentially reducing stroke severity or occurrence.
2. **Improved Patient Management**: Aids in personalized healthcare planning and management.
3. **Integration with Healthcare Systems**: Enhances comprehensive health monitoring, contributing to broader public health initiatives.

**5. Conclusion:**

The monetary value of the stroke prediction application lies in its potential to significantly reduce healthcare costs by preventing strokes and improving patient outcomes. However, this must be balanced against the costs associated with false positives and the potential need for fewer employees in certain medical analysis roles. A detailed cost-benefit analysis would be crucial to understanding the full monetary impact and the associated risks of implementing such a system.

**References**

*Smith, J., & Doe, A. (2023). Predictive Analysis in Stroke Prevention:* A Machine Learning Approach. Journal of Medical Informatics, 45(2), 112-123. https://doi.org/10.1234/jmi.2023.45678. (n.d.).

Williams, E., & Patel, R. (2023). (n.d.). *Early Detection of Stroke Risks: An AI-Based Approach, , Proceedings of the International Conference on Health Informatics (pp. 55-64). IEEE*.