A MIXED INTEGER LINEAR PROGRAMMING MODEL FOR RISK SCORE DEVELOPMENT IN HEALTHCARE

James McKenna; Joseph Agor, PhD
Oregon State University, School of Mechanical, Industrial, and Manufacturing Engineering

BACKGROUND
Electronic Health Record (EHR) data can be used to quantify the relationship between patient attributes and the progression of acute medical conditions. Risk scoring systems can dynamically assess the status of a patient and their risk of further degeneration leading to providers giving preventative care before the onset of a more serious condition. Panels of experts have developed risk scores to assist decision makers at the bedside. Current methods for construction of these scoring systems include statistical modeling and machine learning. With the acceleration of computational power, mixed-integer optimization models have been shown to be an effective way to learn from large data sets through the selection of key features and point values of a risk score.

METHODS
We have created a Mixed Integer Linear Programming (MILP) formulation for the development of a risk scoring system to identify a patient's potential of a specific adverse outcome. Given a set of patient features, our model assigns integer point values representing how much each feature contributes to the risk of the outcome. Current approaches in score development do not attempt to decide upon cutoff values for risk stratification. Instead, cutoff values are determined after the score is developed. In our model we include a constraint to choose a cutoff value for stratifying high and low risk patients (e.g. if the score is greater and 5, then there is a “high risk”). The objective is to maximize the balanced accuracy by correctly placing each observation in a high or low risk scoring interval. Value is gained by increasing the rate of true positives and true negatives given the ratio of actual positive and negative outcomes. A full formulation of the model is given below.

MODEL FORMULATION
We consider an array of data where \( I \) is the number of observations and \( F \) is the number of features. A vector of point values is denoted by \( \lambda \) and is constrained to be an integer with a maximum value of \( P \).

For each observation \( i \) we must bound the vector product of \( \lambda_i \) between the upper and lower bound of a scoring interval which is decided by the optimization model.

The objective is to find the optimal values of \( \lambda \) that maximize the proportion of correctly classified patients. Our objective considers data imbalances by giving equal value to the proportion of correctly identified positive and negative outcomes.

A full formulation of the model is given below.

\[
\begin{align*}
\max & \quad \sum_{i=1}^{N} \sum_{f=1}^{F} x_{if} \lambda_{if} \\
\text{s.t.} & \quad \sum_{i=1}^{N} x_{if} = N, \quad \sum_{i=1}^{N} \lambda_{if} = \lambda_{i}, \quad \lambda_{i} = \max_{j} \lambda_{ij} \\
& \quad 0 \leq \lambda_{ij} \leq M, \quad \forall i \in [1, N], \forall j \in [1, F] \\
& \quad \lambda_{ij} \in \{0, 1\}, \quad \forall i \in [1, N], \forall j \in [1, F] \\
\end{align*}
\]

Features that appear in the original dataset include: sex, polyuria, polydipsia, sudden weight loss, weakness, polyphagia, genitl thrust, visual blurring, itching, irritability, delayed healing, partial paresis, muscle stiffness, alopecia, and obesity.

RESULTS
We compare our results to popular classification algorithms including RiskScore, LASSO, and SVM, all of which use a post hoc cutoff approach. We use statistical testing to compare results from our scoring system to the other classification methods. Our model performs better when the number of variables allowed to be selected is unconstrained. Embedding the cutoffs in the constraints of the optimization problem provides a fully optimized tool for use in numerous applications.

Compared to other classification models, we observe comparable results in the areas sensitivity, specificity, accuracy, and area under the curve (AUC). The computational time is significantly longer, however a feasible solution is still found in acceptable time. We find that there is a need to develop a heuristic solution algorithm for scalability to large datasets.

DISCUSSION
There are some limitations to a MILP approach for risk score development. Using integers for scoring weights provides interpretability of the results, especially for use in a real-world application. However, the addition of features and observational data may increase the computational requirements on an exponential scale. As such, a pragmatic application is to use a smaller subset of data as a training set to determine scoring weights.

SAMPLE OUTPUT
The following is an example of output using the Sylhet Diabetes dataset. We have constrained the model to only select 5 features to provide a compact risk score.

<table>
<thead>
<tr>
<th>Polypuria</th>
<th>Polydipsia</th>
<th>Sudden Weight Loss</th>
<th>Genital Thrush</th>
<th>Irritability</th>
</tr>
</thead>
<tbody>
<tr>
<td>+4</td>
<td>+2</td>
<td>+1</td>
<td>+3</td>
<td>+3</td>
</tr>
</tbody>
</table>

Score = 13

Score ≥ 5: High Risk
Score < 5: Low Risk

SUMMARY OF OUR SOLUTION
We have developed a classification model which uses medical feature data to construct a risk score. While our investigation is ongoing, we have had success in prediction of risk for negative outcomes such as:

1. Preventable hospital admissions due to heart failure
2. Early-stage diabetes detection given a list of symptoms
3. Influenza resistance to immunizations based on comparisons of antigenic distance

We categorized the development of our model into these 5 steps:

• Data preparation and pre-processing includes the selection of high-quality classification data sets. We convert our feature data to binary inputs for our model to solve.

• Formulation and linearization of the model to ensure that an optimal solution to our problem is found. We allow the user to specify parameters such as the total number features to be selected for a score and the largest allowable point value for any of the selected features. By linearizing the model we ensure that the formulation is a valid MILP.

• Score computation with mixed integer programming that produces interpretable risk scores for conditions and adverse outcomes. We use Gurobi optimization software to solve the problem to optimality.

• Statistical hypotheses and testing to compare our results to existing attempts at developing risk scores. We perform a 10-fold cross-validation and split our data sets into 80/20 training and testing subsets to obtain summary statistics. We compare model validation metrics such as sensitivity, specificity, accuracy, F1 score, and area under the receiver operator curve (ROC).

• Development of a heuristic algorithm to arrive at a solution in less time. Using integer variables in mathematical optimization is challenging from a computational perspective. Although a heuristic will provide a sub-optimal solution, the trade-off is a much quicker solution time with a minimal effect on the accuracy of the outcome.