Problem Setting

Your team has been hired as consultants to help a manufacturer understand and address quality issues with a manufacturing process. The manufacturer produces a good that is normally sold at a price of $100. However, the manufacturing process is imperfect and may produce goods with various defects. The manufacturer employs both automated and manual quality inspections of products coming out of the manufacturing process to classify products as "Good" or as belonging to one of three defect classes:

1. Discount - products classified as "Discount" have some aesthetic blemish that prevents them from being sold at full price. Instead, these products are sold at a reduced price of $75 dollars in "discount" outlets.
2. Repair - products classified as "Repair" are still eligible to be sold at discount outlets with some additional effort required to repair some quality issue. The costs of repair can vary, but for the purposes of this analysis, assume that the cost is $5.
3. Scrap - products classified as "Scrap" have some defect that is unrepairable. These products are discarded.

For purposes of cost accounting, assume that the cost to produce a single non-defective product is $50. Also, assume that the firm has done solid marketing research and estimates the annual demand to be 846,312 (assume this is stationary). Finally, assume that any discount products that are produced can be sold and that the firm has the capacity to produce 900,000 products each year.

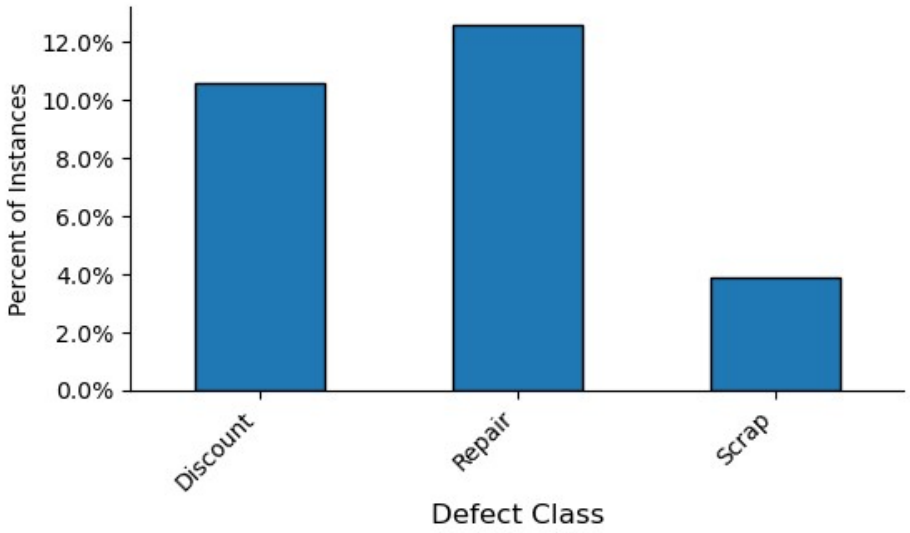
Data

The manufacturer has recently installed 15 sensors at key points in the manufacturing process that collect measurements on various attributes for each product as it goes through the process. They have provided you with a data sample that includes the measurements for each of the sensor readings for 50,000 products. Each row of the dataset provides the readings for a single product and includes additional columns that indicate whether or not the product was classified as belonging to one of the previously described defect classes.

The following figure provides a screenshot of five rows of the data. The columns starting with "sr" (sr01, sr02, …, sr14, sr15) correspond to **s**ensor **r**eadings from the 15 sensors. The columns "Discount," "Repair," and "Scrap" use boolean values to specify whether the product associated with the readings was classified as belonging to each of the defect types. **Note:** a product will be classified as belonging to at most one defect type (or no defect type if the product was good).

sr01	sr02	sr03	sr04	sr05	sr06	sr07	sr08	sr09	sr10	sr11	sr12	sr13	sr14	sr15	Discount	Repair	Scrap
77.54	56.97	15.60	62.04	83.60	26.54	78.82	89.40	8.36	23.89	90.90	30.26	77.51	56.32	101.08	0	0	0
79.52	57.64	16.19	61.90	84.19	23.34	79.42	91.59	8.32	25.07	94.32	26.20	75.10	54.27	94.41	1	0	0
78.81	58.44	17.30	62.89	85.30	26.31	79.72	90.11	8.20	23.51	92.32	28.02	78.24	58.42	99.45	0	0	0
80.44	58.21	17.81	64.28	85.81	27.35	78.30	92.65	6.51	22.77	90.54	28.82	77.89	56.02	98.44	0	0	0
77.72	57.24	16.61	63.21	84.61	25.24	78.02	91.20	8.62	23.81	93.62	28.98	75.03	53.46	95.10	1	0	0

The following bar chart shows the proportion of instances, i.e., rows in the dataset, that are classified as belonging to each defect type.



Goals

Your consultant team has been tasked to use the data to:

- Provide insight regarding which sensor readings are predictive with respect to each defect category. For example, “taken together, the readings from sensors 1 and 12 exhibit are most predictive of whether or not a product will be classified as a “Discount” item”
- Develop a predictive model for each defect class and demonstrate its efficacy in correctly predicting the occurrence of defects.
- Provide ranked recommendations regarding areas of the production process that should be considered for investment to reduce quality issues. For example, “If one area of the production process could be improved, we recommend focusing on the process captured by sensor X. Our justification for this recommendation is Y.”

Deliverable

The terms of your hiring contract specify that you are to prepare a succinct presentation that highlights your findings and recommendations to the management team. This presentation should be no longer than **8 minutes** and address the following points:

- A high-level overview of your process for analyzing the data including the model you selected and any assumptions or special considerations you made in the modeling process,
- Findings regarding the relationship between sensor readings and defect occurrences,
- Your confidence in the identified relationships (based on insights from your predictive modeling work), and
- Your recommendations on areas of investment to reduce defects.