## 

## **Intact Data Science Challenge: Predict Medical Specialty from Medical Notes**

Intact is Canada's largest P&C (property and casualty) insurer. We insure millions of homes, vehicles, and businesses domestically and international through our flagship brand and our subsidiaries.

At Intact, data is our competitive advantage. Data is the key to optimizing our prices, ensuring top customer satisfaction, and to getting the most out of our business processes.

The Data Lab is Intact's in-house Data Science shop, employing hundreds of actuaries, software engineers, and data scientists. Our goal is to develop and deploy data driven solutions to drive change across the organization.

In this challenge, you will have the opportunity to experience what it's like to work on the type of problem we face in Data Lab. Read on to learn about your mission!

### **Challenge Description**

Intact's claims department receives thousands of documents every day. Among these are medical documents, as Intact provides compensation to those who have sustained injuries in automobile accidents. Due to the volume of data received on a daily basis, it is difficult to keep every document neatly organized for easy retrieval. Thus, it would be helpful to have a system that automatically makes sense of the content of the document and can classify it into one of a set number of categories.

In this challenge, you will train a model to do just that: classify a medical transcription to a medical specialty. You are free to train any type of model on the train set. Show any intermediate steps and exploratory data analysis. Use markdown cells to explain any decisions and reasoning.

Your final evaluation will be based on **F1-score** with macro averaging on the test set. However, you will also be judged by the quality of your analysis and the creativity of your solution. Don't be afraid to try new and interesting things, just be sure to document it in a neat way.

Your solution should be self-contained within this notebook.

**Be sure to not leak data from the test set into training. Submissions that do so will be disqualified**.

### **Prize**

* $50 Gift Card for each member in the winning team

### **Submission**

* As with all other challenges, ensure that you have submitted the Google form, the Devpost
* An sample CSV to be submitted to the Google Form can look like this:

| ,Predicted 0,4 1,3 2,10 3,7 4,8 5,2 ... |
| --- |