

CT-223 - Tópicos de Inteligência Artificial

2o. Sem. 2020

(Trabalho em Grupos de Três a Cinco Alunos)

Prof. Paulo André L. Castro

Projeto - Acidente de Trabalho

1. Objetivo

Exercitar e fixar conhecimentos adquiridos sobre Raciocínio Probabilístico, Aprendizado de Máquina e processamento de dados utilizando uma base de dados sobre acidentes de Trabalho em Manufatura de diversos setores.

2. Descrição do Trabalho

2.1. Descrição do Dataset (obtido no Kaggle, o arquivo de dados está disponível no Google classroom)

The database is basically records of accidents from 12 different plants in 03 different countries which every line in the data is an occurrence of an accident.

Columns description

Data: timestamp or time/date information

Countries: which country the accident occurred (anonymized)

Local: the city where the manufacturing plant is located (anonymized)

Industry sector: which sector the plant belongs to (Mining, metals, Others)

Accident level: from I to VI, it registers how severe was the accident (I means not severe ... VI most severe)

Potential Accident Level: Depending on the Accident Level, the database also registers how severe the accident could have been (due to other factors involved in the accident)

Genre: if the person is male or female

Employee or Third Party: if the injured person is an employee or a third party

Critical Risk: some description of the risk involved in the accident

2.2. Classificador Bayes Ingênuo

Utilizando a base de dados fornecida, criar um **classificador baseado em Nãive Bayes** que classifique o nível do acidente dadas as demais informações. Discuta quais variáveis são mais ou menos relevantes para o processo de decisão.

2.3. Classificador Bayesiano

Utilizando a base de dados fornecida, criar um **classificador bayesiano com estrutura determinada a partir dos dados**, teste diferentes parâmetros de treinamento de modo a tentar encontrar um modelo que supere o modelo **Nãive Bayes**. Discuta quais variáveis são mais ou menos relevantes para o processo de decisão

2.4. Análise Comparativa

Compare os dois classificadores utilizando: taxa de acerto, matriz de confusão, erro quadrático médio e estatística kappa utilizando a técnica de 10-folds.

2.5. Pre-processamento do conjunto de dados e Implementação

Faz parte do trabalho fazer o pré-processamento dos dados de modo a permitir seu uso nos algoritmos de aprendizado e frameworks selecionados. Sugerimos o uso do python ou julia para este pré-processamento, para a construção e avaliação dos classificadores sugerimos o scikit-learn ou Weka, mas você pode usar qualquer framework ou linguagem de sua preferência.

3. Material a ser Entregue e Prazo

Material: Relatório e Código

Prazo de Entrega: 30/Novembro/2020

Relatório do Projeto (arquivo em formato pdf até 4 páginas) com:

Título e Nomes dos integrantes do Grupo

1. Resumo (max. 2 parágrafos)

2. Descrição e Resultados Obtidos

2.1. Descrição dos Classificadores e do dataset

2.2. Dados e Resultados da comparação

2.3. Discussão e sugestão de melhorias possíveis para o classificador

2.4. Apresente padrões relevantes existentes nos dados

(Por exemplo, variáveis que podem contribuir para acidentes mais graves. Procure verificar também se existe algum tipo de sazonalidade nos acidentes: meses, dias da semana, etc.)

3. Conclusões: Comentários e sugestões sobre o trabalho (complexidade/facilidade, sugestões, etc.).

4. Implementação: Descrição da implementação, comentários eventualmente necessários para a execução do projeto e Código do Projeto.

Bom Trabalho!
Prof. Paulo André Castro
pauloac@ita.br