

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA
Pós-graduação *Lato Sensu* em Inteligência Artificial e Aprendizado de Máquina

Victor Hugo Simões Pinheiro Filho

**Análise de segurança em atividade offshore: adequação do planejamento com
uso de rede neural**

Macaé
2021

Victor Hugo Simões Pinheiro Filho

**Análise de segurança em atividade offshore: adequação do planejamento com
uso de rede neural**

Trabalho de Conclusão de Curso apresentado
ao Curso de Especialização em Inteligência
Artificial e Aprendizado de Máquina como
requisito parcial à obtenção do título de
especialista.

Macaé

2021

SUMÁRIO

1. Introdução.....	5
1.1. Contextualização.....	5
1.2. O problema proposto.....	5
2. Coleta de Dados	6
3. Processamento/Tratamento de Dados	6
4. Análise e Exploração dos Dados	7
5. Criação de Modelos de Machine Learning	11
6. Apresentação dos Resultados	12
7. Conclusão	13
8. Links	14
9. Glossário.....	15
REFERÊNCIAS.....	16

Lista de Imagens

Figura 1 – Dados sem tratamento do arquivo csv.....	6
Figura 2 – Dados após o pivoteamento do dataset de características.....	7
Figura 3 – Dados sem tratamento do arquivo csv.....	7
Figura 4 – Dados após o pivoteamento do dataset EPI	7
Figura 5 - Mapa da correlação entre características de entrada	8
Figura 6 - Correlação linear entre H2S e Liberação de Tanques.....	8
Figura 7 - Mapa de correlação dos EPI.....	9
Figura 8 - Colunas removidas em EPI	9
Figura 9 - Todas as colunas Óculos	10
Figura 10 - Criação da colunas OCULOS	10
Figura 11 - Merge entre datasets	10
Figura 12 - Correlação linear entre entrada e a saída	11
Figura 13 - Divisão dos dados.....	11
Figura 14 - Modelo Python para previsão	12
Figura 15 - Execução do modelo	12
Figura 16 - Perda x Iterações.....	12
Figura 17 - Acurácia x Iteração	13

1. Introdução

1.1. Contextualização

A indústria petrolífera requer planejamento diário das atividades de campo, desde a correta aquisição de equipamentos para a execução de manutenções preditivas e corretivas, além da análise de segurança da atividade em si.

A manutenção da integridade física do trabalhador é prioridade em uma área que envolve tantos riscos. Para cada trabalho exercido, existem vários elementos envolvidos que caracterizam o grau de exposição ao risco. Por exemplo, uma mesma atividade pode envolver equipamentos pressurizados ou elétricos, instrumentos que causam centelha, área com possibilidade de gás inflamável, apenas para citar o quão necessário se torna o planejamento adequado do serviço a ser executado. O planejador precisa de ferramentas que se baseiem na experiência passada de execução para ajudá-lo no planejamento de novas atividades.

1.2. O problema proposto

Um dos muitos problemas recorrentes do processo de planejamento dessas atividades é a correta alocação de recursos de EPI. Na atividade de planejamento os recursos são alocados e devem estar disponíveis durante a execução da atividade. O planejamento inadequado pode causar indisponibilidade do EPI no momento correto de sua utilização, levando a uma maior insegurança durante a operação.

A atividade está dividida em duas etapas: planejamento e execução. Na primeira etapa, o EPI a ser alocado é informado pelo planejador. Na etapa de execução, o EPI é utilizado pelo operador responsável pela execução da atividade. Naturalmente, podem ocorrer discrepâncias entre o EPI informado pelo planejador e pelo executante.

Uma ferramenta de inteligência artificial poderia informar ao planejador os EPIs necessários para a execução da atividade, baseando-se para isso, no histórico

de execução da atividade. As características do trabalho (Chama aberta, Gás inflamável, H₂S, entre outros) podem constituir fatores importantes na determinação do EPI utilizado.

Os dados analisados foram extraídos de uma unidade marítima de exploração de petróleo durante o ano de 2020.

2. Coleta de Dados

Os dados foram obtidos na base de dados Oracle transacional utilizada pela área operacional e o resultado convertido para arquivo csv. São dois datasets extraídos dessa base: um contendo as características de trabalho e o outro contendo os EPIs utilizados para cada permissão de trabalho.

caracP50.csv

NUMEROPT	CARACTERISTICA
----------	----------------

epiP50.csv

NUMEROPT	EPI
----------	-----

3. Processamento/Tratamento de Dados

Na primeira etapa de transformação dos dados, foi realizado o pivoteamento dos dados indicando para cada permissão de trabalho, se determinada característica está ou não presente. Por exemplo, se a permissão 10/2020 possui a característica Pressurizado, deve-se assinalar no dataset o valor 1, para as outras características não presentes, o valor 0. A codificação utilizada é one-hot encode.

NUMEROPT	CARACTERISTICA	count
1/2020	Pressurizado	1
16/2020	Pressurizado	1
24/2020	Pressurizado	1
27/2020	Risco de Perda de Produção	1
37/2020	Pressurizado	1

Figura 1 – Dados sem tratamento do arquivo csv

```

caracteristicasfinal = caracteristicas.pivot_table(index=["NUMEROPT"],
columns=['CARACTERISTICA'], values='count').fillna(0)
caracteristicasfinal.head()

```

CARACTERISTICA	Acesso e trabalho em casas de bombas	Altura - Acesso por Cordas	Andaime - Pré-Montagem/Desmontagem	Andaime - Pós-Montagem/Utilização	Chama Aberta	Corte e Solda	Elétrico	Esmerilhadeira e Lixadeira
NUMEROPT								
1/2020	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10/2020	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
100/2020	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
10000/2020	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0
10001/2020	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0

Figura 2 – Dados após o pivoteamento do dataset de características

O dataset com as características será a entrada do modelo. Ele possui 6763 registros e 33 colunas. Para o dataset de EPI, foi realizada a mesma abordagem.

NUMEROPT	EPI	count
4/2020	ÓCULOS SEGURANÇA CONTRA IMPACTO	1
6/2020	EPI's OBRIGATÓRIOS (CAPACETE, BOTA, PROT. AURI...	1
8/2020	ÓCULOS PROTETOR FACIAL	1
8/2020	EPI's OBRIGATÓRIOS (CAPACETE, BOTA, PROT. AURI...	1
11/2020	EPI's OBRIGATÓRIOS (CAPACETE, BOTA, PROT. AURI...	1

Figura 3 – Dados sem tratamento do arquivo csv

EPI	ABAFADOR	AR MANDADO	AR AVENTAL	AVENTAL DE RASPA	BALACLAVA	BLUSÃO DE RASPA	BOTA DE CANO ALTO	BOTA DE COURO COM BIQUETRA COMPOSITE	BOTA DE COURO COM BIQUETRA DE AÇO	BOTA DE COURO COM BIQUETRA PLÁSTICA	BOTA DE COURO SEM BIQUETRA	BOTA DE PVC	BOTA DE PVC CANO LONGO	BOTA DE SEGURANÇA PARA ELETRICISTA
NUMEROPT														
10/2020	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
100/2020	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1000/2020	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10000/2020	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10001/2020	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Figura 4 – Dados após o pivoteamento do dataset EPI

O dataset com os EPIs será a saída do modelo. Ele possui 9517 registros e 120 colunas. Existem muitas colunas de saída, que em um modelo com grande quantidade de registros podem demandar razoável esforço computacional. É necessário encontrar formas de diminuir a quantidade de colunas.

4. Análise e Exploração dos Dados

A primeira etapa é utilizar tentar excluir colunas com utilização de correlação linear. Se as colunas de entrada possuem alta correlação, pode-se excluir algumas. Com o uso do seaborn, obtivemos o mapa de correlação abaixo.

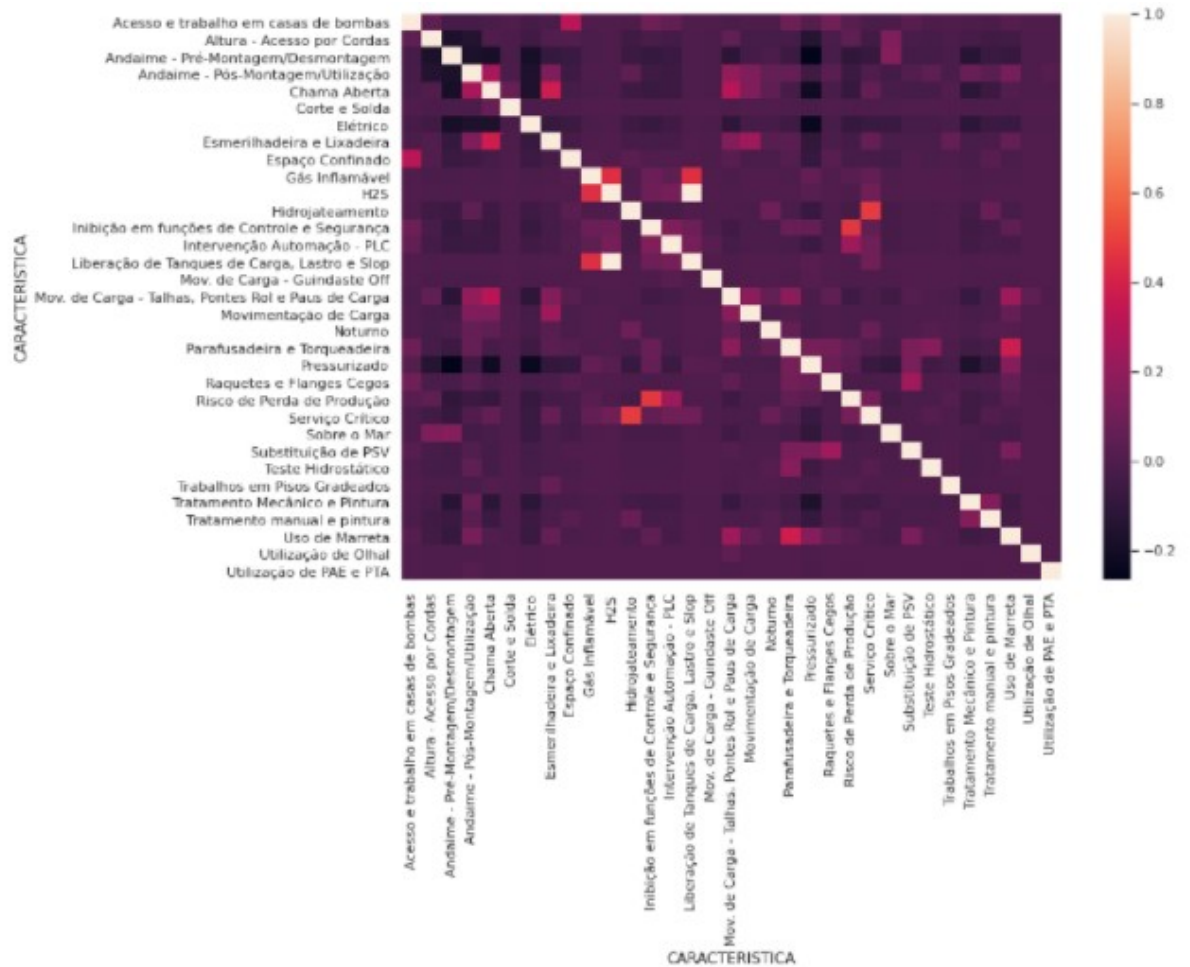


Figura 5 - Mapa da correlação entre características de entrada

Para remover colunas que possuem correlação acima de 80%, obteve-se apenas a exclusão de uma coluna.

```
Liberação de Tanques de Carga, Lastro e Slop | H2S | 1.0
(6763, 32)
```

Figura 6 - Correlação linear entre H2S e Liberação de Tanques

O mapa de correlação mostrou que essas duas características de trabalho sempre aparecem juntas, podemos, portanto, excluir uma delas sem prejuízo da análise. O dataset possui agora 32 colunas.

O mesmo procedimento foi adotado para o dataset de EPI, obtendo o mapa de correlação abaixo.

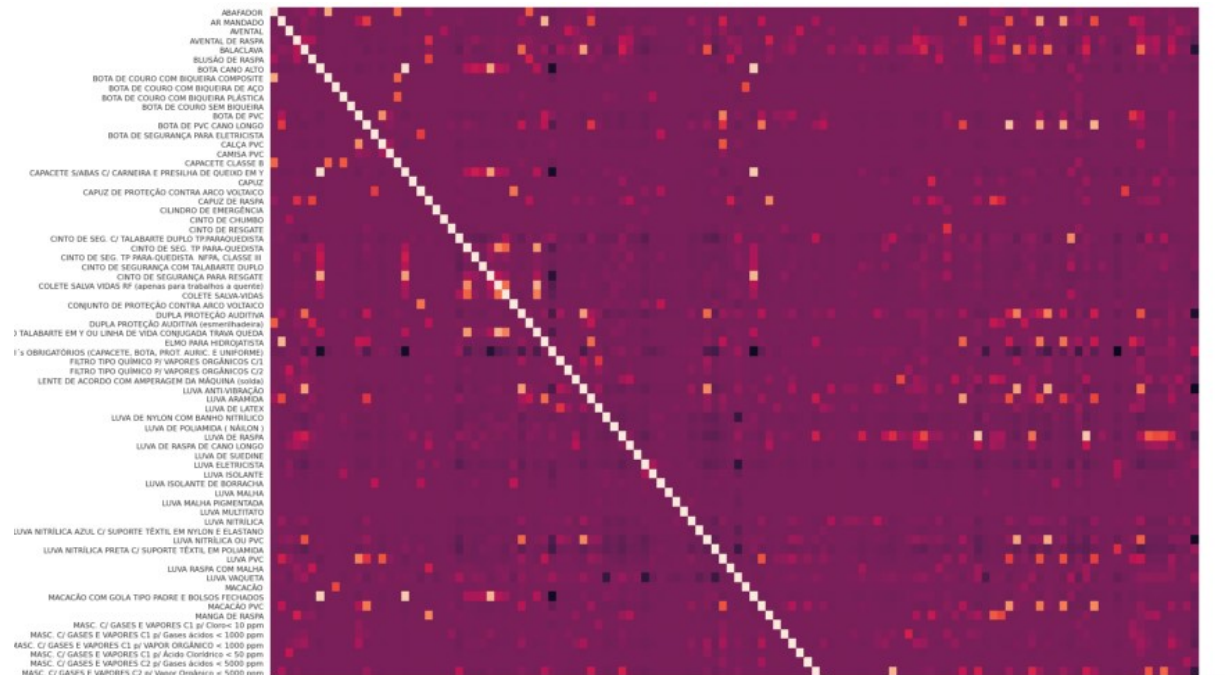


Figura 7 - Mapa de correlação dos EPI

Para remover colunas que possuem correlação acima de 80%, houve redução de 120 para 110 colunas.

CAPACETE S/ABAS C/ CARNEIRA E PRESILHA DE QUEIXO EM Y | BOTA CANO ALTO | 0.95
 ELMO PARA HIDROJATISTA | AR MANDADO | 0.8
 MACACÃO COM GOLA TIPO PADRE E BOLSOS FECHADOS | BOTA CANO ALTO | 0.89
 MACACÃO COM GOLA TIPO PADRE E BOLSOS FECHADOS | CAPACETE S/ABAS C/ CARNEIRA E PRESILHA DE QUEIXO EM Y | 0.88
 MÁSCARA SOLDADOR | LUVA DE RASPA | 0.85
 PEÇA FACIAL COMPLETA COM INDUÇÃO DE AR | BOTA DE PVC CANO LONGO | 0.84
 PROTEÇÃO TIPO CASCO DE TATU | PEÇA FACIAL COMPLETA COM INDUÇÃO DE AR | 0.89
 RESPIRADOR COM FILTRO PARA GASES ÁCIDOS E VAPORES ORGÂNICOS | PEÇA SEMI-FACIAL FILTRANTE 2 | 0.92
 SOBRE BOTA DE ARAMIDA | PEÇA FACIAL COMPLETA COM INDUÇÃO DE AR | 0.9
 SOBRE BOTA DE ARAMIDA | PROTEÇÃO TIPO CASCO DE TATU | 0.99
 VESTIM. COMPLETA DE RASPA | LUVA DE RASPA | 0.87
 VESTIM. COMPLETA DE RASPA | MÁSCARA SOLDADOR | 0.9
 VESTIM. COMPLETA PVC | PEÇA FACIAL COMPLETA COM INDUÇÃO DE AR | 0.95
 VESTIM. COMPLETA PVC | PROTEÇÃO TIPO CASCO DE TATU | 0.85
 VESTIM. COMPLETA PVC | SOBRE BOTA DE ARAMIDA | 0.85
 (9517, 110)

Figura 8 - Colunas removidas em EPI

A redução é pequena, o que ainda não ajuda na redução do tempo computacional para grandes quantidades de registros. Desta forma, foi adotada a forma de agrupamento semântico. Todos os EPIs do tipo ÓCULOS serão agrupados em uma única coluna chamada OCULOS. O primeiro passo é selecionar apenas as colunas com esse tipo de EPI.

```
oculosdf = epicorr.filter(like='ÓCULO')
oculosdf = oculosdf.iloc[:, 0:6]
oculosdf.head()
```

EPI	ÓCULOS AMPLA VISÃO	ÓCULOS AMPLA VISÃO (esmerilhadeira)	ÓCULOS MAÇARIQUEIRO	ÓCULOS MAÇARIQUEIRO (oxicorte)	ÓCULOS PROTETOR FACIAL	ÓCULOS SEG. CONTRA POEIRA
NUMEROPT						
10/2020	0.0	0.0	0.0	0.0	0.0	0.0
100/2020	0.0	0.0	0.0	0.0	0.0	0.0
1000/2020	0.0	0.0	0.0	0.0	0.0	0.0
10000/2020	0.0	0.0	0.0	0.0	0.0	0.0
10001/2020	0.0	0.0	0.0	0.0	0.0	0.0

Figura 9 - Todas as colunas Óculos

O segundo passo é agrupar essas colunas em apenas uma.

```
oculosdf['OCULOS'] = oculosdf[:].max(axis=1)
oculosdf['OCULOS'].value_counts()
saida = oculosdf.filter(like='OCULO')
saida.head()
```

EPI	OCULOS
NUMEROPT	
10/2020	0.0
100/2020	0.0
1000/2020	0.0
10000/2020	0.0
10001/2020	0.0

Figura 10 - Criação da colunas OCULOS

Até este momento, ainda temos dois datasets separados, com quantidade de registros diferentes. Devemos fazer o merge entre eles, utilizando a coluna NUMEROPT como a coluna de junção. Agora temos um dataset com 6299 registros, sendo 32 colunas de entrada e a coluna OCULOS de saída. Esta coluna indica se algum EPI de tipo óculos foi informado.

```
result = pd.merge(caracteristicacorr, saida, on="NUMEROPT")
result.shape

(6299, 33)
```

Figura 11 - Merge entre datasets

Após a realização do merge dos dois datasets, executamos o algoritmo de correlação linear. Constata-se a correlação moderada entre as características de

chama aberta e tratamento mecânico e pintura com o EPI óculos. Quando alguma dessas características aparece, aumenta a chance da necessidade do uso de óculos de segurança na permissão do trabalho.



Figura 12 - Correlação linear entre entrada e a saída

5. Criação de Modelos de Machine Learning

Dadas as características de trabalho, é necessário prever se o EPI óculos estará ou não presente. É um problema de classificação binária. Para esta predição, utilizou-se uma rede neural construída com Keras. A rede neural possui 3 camadas densas com 3 neurônios cada e uma camada de saída com o resultado da predição.

Para a função de perda, utilizamos a entropia cruzada(binary-cross-entropy), que é mais indicada para classificadores binários.

Os dados para treinamento e teste foram divididos em proporção 80% e 20% respectivamente. Os dados de treinamento foram separados em 10% para validação. A construção do modelo encontra-se no código abaixo:

```
from sklearn.model_selection import train_test_split
#Create train and test dataset with an 80:20 split

x_train, x_test, y_train, y_test = train_test_split(entrada,result['OCULOS'],test_size=0.2)
#Further divide training dataset into train and validation dataset with an 90:10 split
x_train, x_val, y_train, y_val = train_test_split(x_train,y_train,test_size=0.1)
#Check the sizes of all newly created datasets
print("Shape of x_train:",x_train.shape)
print("Shape of x_val:",x_val.shape)
print("Shape of x_test:",x_test.shape)
print("Shape of y_train:",y_train.shape)
print("Shape of y_val:",y_val.shape)
print("Shape of y_test:",y_test.shape)
```

Figura 13 - Divisão dos dados

```

from keras.models import Sequential
from keras.layers import Dense, Dropout
from keras.callbacks import History
from keras.layers.core import Dropout, Dense

history = History()

model = Sequential()
model.add(Dense(3, activation='relu', input_shape=(32,)))
model.add(Dense(3, activation='relu'))
model.add(Dense(3, activation='relu'))
model.add(Dense(1, activation='sigmoid'))

#Configure the model
model.compile(loss='binary_crossentropy',
              optimizer='adam',
              metrics=['accuracy'])
#Train the model
model.fit(x_train.values, y_train.values, validation_data=(x_val, y_val), epochs=5, batch_size=32, callbacks=[history])

```

Figura 14 - Modelo Python para previsão

6. Apresentação dos Resultados

Após executar em 5 iterações, o modelo apresentou acurácia de 92%.

```

Epoch 1/5
142/142 [=====] - 1s 2ms/step - loss: 0.6562 - accuracy: 0.7509 - val_loss: 0.5406 - val_accuracy: 0.8770
Epoch 2/5
142/142 [=====] - 0s 1ms/step - loss: 0.4975 - accuracy: 0.8598 - val_loss: 0.3210 - val_accuracy: 0.9107
Epoch 3/5
142/142 [=====] - 0s 1ms/step - loss: 0.2965 - accuracy: 0.8966 - val_loss: 0.2307 - val_accuracy: 0.9325
Epoch 4/5
142/142 [=====] - 0s 1ms/step - loss: 0.2383 - accuracy: 0.9193 - val_loss: 0.2177 - val_accuracy: 0.9286
Epoch 5/5
142/142 [=====] - 0s 1ms/step - loss: 0.2063 - accuracy: 0.9273 - val_loss: 0.2155 - val_accuracy: 0.9345

```

Figura 15 - Execução do modelo

Para analisar se o modelo está generalizando bem, precisamos verificar os gráficos de perda e acurácia.

A perda cai consistentemente para os dados de treinamento e validação, não apresentando discrepância entre ambos.



Figura 16 - Perda x Iterações

A acurácia cresce consistentemente para os dados de treinamento e validação em cada iteração. Em situações em que se observa um grande gap entre dados de treinamento e validação seria um indicador de overfitting do modelo, o que parece não ocorrer.

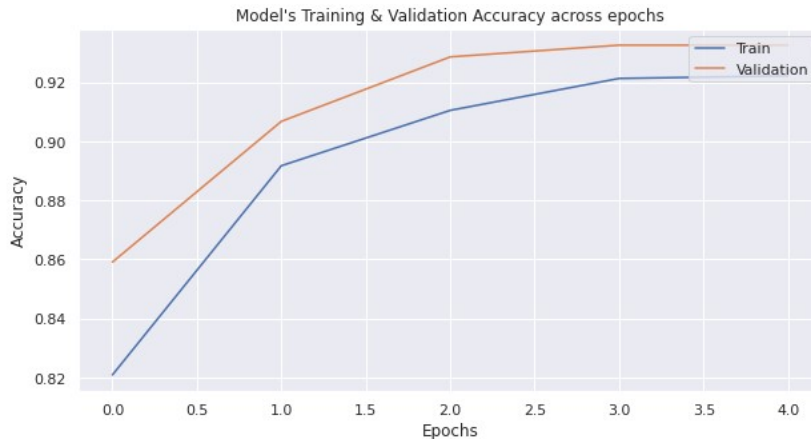


Figura 17 - Acurácia x Iteração

Podemos concluir que para a unidade marítima analisada, podemos informar com precisão de 92% ao planejador, que o EPI de tipo óculos é necessário ou não para a atividade planejada. O sistema poderá fazer uma recomendação acurada caso perceba a discrepância entre o planejamento e a execução da atividade.

7. Conclusão

As características de trabalho constituem fator importante na identificação do risco de uma atividade, porém não são os únicos. Há outros dados das permissões de trabalho que devem ser considerados e que não fizeram parte do escopo deste trabalho.

Com o uso de correlação linear, foi possível identificar relação entre as características e os EPIs utilizados, e com uso de rede neural, pudemos fazer previsão sobre EPI necessário à atividade baseado na experiência, ao invés da intuição do planejador.

A mesma abordagem será utilizada para identificar outras variáveis na identificação e mitigação dos riscos.

8. Links

Material do TCC no github: <https://github.com/victorpucminas/tcc>

9. Glossário

NUMEROPT: Número da Permissão de Trabalho.

PT: Permissão de Trabalho. É a execução da atividade planejada.

Características de Trabalho: Elementos que compõem o risco de uma atividade dentro de uma área.

H₂S: Sulfeto de hidrogênio. Gás inodoro em alta concentração.

EPI: Equipamento de proteção individual. Vestimentas utilizadas pelo trabalhador nas áreas operacionais. Pode ser óculos, luvas, máscaras, etc.

Chama Aberta: Toda atividade que envolva uso de maçarico ou fonte ignitória.

REFERÊNCIAS

Moolayil, Jojo. **Learn Keras for deep neural networks.**

Loy, James. **Neural Network Projects with Python.**

Documentação Pandas: <https://pandas.pydata.org/docs/>