

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS
NÚCLEO DE EDUCAÇÃO A DISTÂNCIA
Pós-graduação *Lato Sensu* em Ciência de Dados e Big Data

Luis Carlos Batista Santos

Machine Learning aplicado aos dados de
Ocorrências Aeronáuticas no Brasil

Brasília, DF
2020

Luis Carlos Batista Santos

**Machine Learning aplicado aos dados de
Ocorrências Aeronáuticas no Brasil**

Trabalho de Conclusão de Curso apresentado
ao Curso de Especialização em Ciência de
Dados e Big Data como requisito parcial à
obtenção do título de especialista.

Brasília, DF

2020

SUMÁRIO

1. Introdução.....	4
1.1. Contextualização	4
1.2. O problema proposto	6
2. Coleta de Dados	8
3. Processamento/Tratamento de Dados	11
4. Análise e Exploração dos Dados	16
5. Criação de Modelos de Machine Learning	21
6. Apresentação dos Resultados	27
7. Links	34
REFERÊNCIAS.....	36

1. Introdução

1.1. Contextualização

A Aviação Civil tem um relevante papel na atividade socioeconômica no Brasil. Com o crescimento deste setor, a segurança de voo é primordial para o equilíbrio na operação aeronáutica do País. E dentro deste cenário destacam-se os operadores aeronáuticos que trabalham para regular, fiscalizar, investigar, gerenciar as atividades no setor aéreo: Agência Nacional de Aviação Civil (ANAC), Comando da Aeronáutica (COMAER), Departamento de Controle do Espaço Aéreo (DECEA), Operadores Aeronáuticos e Operadores de Infraestrutura Aeroportuária.

Dessa forma, as organizações tem o desafio de criar e manter uma cultura de segurança operacional a fim de controle e minimização do risco que deve estar no topo da organização, e desta forma gerar relatos de eventos.

As ocorrências desses eventos em grande volume são notificadas ao Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA), que é o órgão pertencente ao Comando da Aeronáutica (COMAER), responsável legal para investigar ocorrências aeronáuticas e por todo o processo gerador de informações coletadas. À vista disso, a coleta de dados visa identificar padrões entre conjunto de variáveis para melhor tomada de decisão e mitigar os riscos.

O CENIPA é o órgão central do SIPAER, que tem por finalidade planejar, gerenciar, controlar e executar as atividades correlatas a prevenção e investigação de acidentes aeronáuticos. O Manual do Comando da Aeronáutica (MCA 3-3) relata que tal Norma tem por finalidade estabelecer protocolos, responsabilidades e atribuições para o planejamento e a execução das atividades básicas da prevenção de acidentes aeronáuticos, permitindo que os Elos do Sistema de Investigação e Prevenção de Acidentes Aeronáuticos (SIPAER) desenvolvam tais atividades de acordo com as características das missões e as especificidades de suas organizações, visando à melhoria da Segurança de Voo das operações aéreas no Brasil (CENIPA, 2012). Ainda neste contexto, o Elo SIPAER pode ser um órgão, setor ou cargo, dentro da estrutura das organizações, com responsabilidade para ministrar assuntos de segurança de voo no âmbito SIPAER.

O NSCA 3-13/2017 (Protocolos de Investigação de Ocorrências Aeronáuticas da Aviação Civil conduzidas pelo Estado Brasileiro) define ocorrência Aeronáutica como qualquer evento envolvendo aeronave que possa ser classificado como acidente aeronáutico, incidente aeronáutico grave, incidente aeronáutico ou ocorrência de solo (somente na aviação militar), permitindo ao SIPAER a adoção dos procedimentos pertinentes.

E de acordo com o Manual de Investigação do SIPAER (MCA 3-6):

Acidente Aeronáutico

Toda ocorrência aeronáutica relacionada à operação de uma aeronave tripulada, havida entre o momento em que uma pessoa nela embarca com a intenção de realizar um voo até o momento em que todas as pessoas tenham dela desembarcado ou; no caso de uma aeronave não tripulada, toda ocorrência havida entre o momento que a aeronave está pronta para se movimentar, com a intenção de voo, até a sua parada total pelo término do voo, e seu sistema de propulsão tenha sido desligado e, durante os quais, pelo menos uma das situações abaixo ocorra:

a) uma pessoa sofra lesão grave ou venha a falecer como resultado de:

- estar na aeronave;
- ter contato direto com qualquer parte da aeronave, incluindo aquelas que dela tenham se desprendido; ou
- ser submetida à exposição direta do sopro de hélice, de rotor ou de escapamento de jato, ou às suas consequências.

b) a aeronave tenha falha estrutural ou dano que:

- afete a resistência estrutural, o seu desempenho ou as suas características de voo; ou
- normalmente exija a realização de grande reparo ou a substituição do componente afetado.

c) a aeronave seja considerada desaparecida ou esteja em local inacessível.

Incidente Aeronáutico

Uma ocorrência aeronáutica, não classificada como um acidente, associada à operação de uma aeronave, que afete ou possa afetar a segurança da operação.

Incidente Aeronáutico Grave

Incidente aeronáutico envolvendo circunstâncias que indiquem que houve elevado risco de acidente relacionado à operação de uma aeronave que, no caso de aeronave tripulada, ocorre entre o momento em que uma pessoa nela embarca, com a intenção de realizar um voo, até o momento em que todas as pessoas tenham dela desembarcado; ou, no caso de uma aeronave não tripulada, ocorre entre o momento

em que a aeronave está pronta para se movimentar, com a intenção de voo, até a sua parada total pelo término do voo, e seu sistema de propulsão tenha sido desligado.

NOTA 1 - A diferença entre o incidente grave e o acidente está apenas nas consequências.

1.2. O problema proposto

A Aviação civil brasileira é formada por aeronaves de uso comercial ou privado. Sendo categorizadas em três grupos:

- ✓ Aviação comercial: todo serviço de transporte aéreo público prestado a passageiro e carga, pode ser regular ou não regular.
- ✓ Aviação geral: todas as operações de aviação civil que não sejam serviços aéreos regulares nem operações não regulares de transporte aéreo por remuneração ou arrendamento. São aviões de propriedade particular, serviços aéreos especializados, aeronaves de instrução, etc.
- ✓ Aviação experimental: são aeronaves com fins experimentais e desportivas como planador, ultraleve, girocôptero, balão e helicóptero ultraleve.

Neste enquadramento, o total de aeronaves registradas no Registro Aeronáutico Brasileiro (RAB), até junho de 2019, foi de 22219, sendo que 65%, 26% e 9% são respectivamente da aviação geral, aviação experimental e aviação comercial (ANAC, 2019) e o número de pousos e decolagens registrados na aviação civil foram mais de 2,4 milhões (Anuário Brasileiro da Aviação Civil, 2019).

Os operadores aeronáuticos brasileiros, preocupados em promover um serviço seguro à sociedade, adotou o SGSO (Sistema de Gerenciamento de Segurança Operacional) que foi implementado pela ICAO (Organização Internacional da Aviação Civil) em 2001 por toda aviação mundial. O SGSO tem como processo chaves: identificação de perigos, reporte de eventos de segurança operacional, gerenciamento de risco, medição de desempenho e garantia de qualidade (ANAC).

Por definição, SGSO é um conjunto de ferramentas gerenciais e métodos organizacionais para apoiar as decisões a serem tomadas por um provedor de serviços da aviação civil em relação ao risco de suas atividades diárias (ANAC).

Os métodos de gerenciamento da segurança operacional, conforme ICAO, são categorizadas como:

- ✓ Reativo: análise de ocorrências, determinação de fatores contribuintes e decisões sobre discrepâncias relacionadas a riscos.
- ✓ Proativo: sistema mandatórios e voluntários de relatórios, auditorias de segurança operacional, e pesquisas.
- ✓ Preditivo: sistema de relatórios confidenciais, análise de dados de voo, e monitoramento das operações normais.

Os modelos são excelentes ferramentas para ajudar a prever o futuro. E são usados na previsão de segurança operacional da aviação. A Mineração de Dados, um método analítico, procura fazer maior uso das bases de dados para aprender mais sobre o problema ou questões em suas mãos (STOLZER, HALFORD & GOGLIA, 2011).

Através da Mineração de Dados procura-se descoberta de padrões ocultos nos dados através do aprendizado de máquina. Enquanto a mineração de dados procura revelar conexões e associações previamente desconhecidos nos dados, a aprendizagem de máquina foca na análise de dados e sua compreensão ensinando um programa de computador a fazer o trabalho (Análise Preditiva para leigos, 2019).

Conforme, Foster Provost & Tom Fawcett, o campo da Mineração de Dados (KDD: knowledge-discovery in databases/ Descoberta de Conhecimento e Mineração de Dados) começou como uma ramificação do Aprendizado de Máquina, e permanecem intimamente ligados.

Uma análise preditiva na base de dados de ocorrências aeronáuticas da aviação civil brasileira tem como objetivo identificar insights nos registros de acidentes, incidentes graves e incidentes ocorridos na aviação comercial, geral e experimental.

Dessa forma, este trabalho visa aplicar um modelo preditivo a partir de dados de ocorrências aeronáuticas na aviação civil brasileira registrados entre 2009 a maio de 2019. O QlikView Versão 11.20.13206.0 SR13 64-bit Edition (x64) é usado para o relacionamento das quatro tabelas e construção de uma tabela única com as principais feature. O Excel (Microsoft Office Professional Plus 2016) é usado para comportar os dados em xlsx. O Orange Versão: 3.24.1 será utilizado para desenvolvimento de modelos preditivos de classificação ou indução supervisionada. A classificação tem como objetivo analisar os dados históricos numa base de dados para depois gerar um modelo capaz de prever o comportamento futuro.

2. Coleta de Dados

A base de dados está disponível no Portal de Dados Abertos do Governo Federal no endereço: <http://www.dados.gov.br/dataset/ocorrencias-aeronauticas-da-aviacao-civil-brasileira>. E foi acessada em 20 de janeiro de 2020. Os dados estão disponíveis em tabelas (arquivos) no formato csv (comma-separated-values), valores separados por vírgula:

oco.csv- informações sobre as ocorrências.

anv.csv- informações sobre as aeronaves envolvidas nas ocorrências.

ftc.csv - informações sobre os fatores contribuintes nas ocorrências.

rec.csv – informações sobre as recomendações de segurança geradas nas ocorrências.

A tabela abaixo mostra as variáveis da tabela ocorrência. Esta tabela é composta por 22 variáveis conforme sua descrição e tipo. Todas as ocorrências aeronáuticas foram no Brasil.

Tabela 1

tabela_oco	descricao	tipo
codigo_ocorrencia	id	númeroico
ocorrencia_classificacao	ACIDENTE	Categórica nominal
	INCIDENTE	
	INCIDENTE GRAVE	
ocorrencia_tipo	diversos=71 níveis	Categórica nominal
ocorrencia_tipo_categoria	diversos=31 níveis	Categórica nominal
ocorrencia_tipo_icao	diversos=29 níveis	Categórica nominal
ocorrencia_latitude	Numero=2249	Categórica nominal
ocorrencia_longitude	Numero=2259	Categórica nominal
ocorrencia_cidade	diversos=1050	Categórica nominal

ocorrencia_uf	diversos=28	Categórica nominal
ocorrencia_pais	BRASIL	Categórica nominal
ocorrencia_aerodromo	designatico OACI=480	Categórica nominal
ocorrencia_dia	data	data
ocorrencia_horario	hora	hora
investigacao_aeronave_liberada	***	Categórica nominal
	NÃO	
	NULL	
	SIM	
investigacao_status	ATIVA	Categórica nominal
	FINALIZADA	
divulgacao_relatorio_numero	diversos=1786	string
divulgacao_relatorio_publicado	NÃO	Categórica nominal
	SIM	
divulgacao_dia_publicacao	data	data
total_recomendacoes	Numero	numerico
total_aeronaves_envolvidas	numero	numerico
ocorrencia_saida_pista	NÃO	Categórica nominal
	SIM	
ocorrencia_dia_extracao	data	data

A tabela abaixo mostra as variáveis da tabela aeronave. Esta tabela é composta por 25 variáveis conforme sua descrição e tipo.

Tabela 2

tabela_anv	descricao	tipo
codigo_ocorrencia	id	númeroico
aeronave_matricula	varias matriculas=3998	string
aeronave_operador_categoria	AERoclube	Categórica nominal
	ESCOLA DE AVIAÇÃO	
	FABRICANTE AERONÁUTICO	
	OPERADOR DE AERONAVE	
	OPERADOR PARTICULAR	
	ÓRGÃO PÚBLICO	
aeronave_tipo_veiculo	ANFÍBIO	Categórica nominal
	AVIÃO	
	BALÃO	
	DIRIGÍVEL	
	HELICÓPTERO	
	HIDROAVIÃO	

	INDETERMINADA	
	PLANADOR	
	TRIKE	
	ULTRALEVE	
aeronave_fabricante	diversos=231	string
aeronave_modelo	diversos=783	string
aeronave_tipo_icao	diversos=237	string
aeronave_motor_tipo	***	Categórica nominal
	JATO	
	PISTÃO	
	SEM TRAÇÃO	
	TURBOEIXO	
	TURBOÉLICE	
aeronave_motor_quantidade	***	Categórica nominal
	BIMOTOR	
	MONOMOTOR	
	QUADRMOTOR	
	SEM TRAÇÃO	
	TRIMOTOR	
aeronave_pmd	numero	numerico
aeronave_pmd_categoria	***	Categórica ordinal
	ACIMA DE 5700 KG	
	ATÉ 2250 KG	
	ENTRE 2250 E 5700 KG	
aeronave_assentos	numero	numerico
aeronave_ano_fabricacao	ano	data
aeronave_pais_fabricante	país=17	Categórica nominal
aeronave_pais_registro	país=23	Categórica nominal
aeronave_registro_categoria	diversos=33	Categórica nominal
aeronave_registro_segmento	***	Categórica nominal
	ADMINISTRAÇÃO DIRETA	
	ADMINISTRAÇÃO INDIRETA	
	AGRÍCOLA	
	ESPECIALIZADA	
	EXPERIMENTAL	
	HISTÓRICA	
	INSTRUÇÃO	
	MÚLTIPLA	
	NÃO REGULAR	
	PARTICULAR	
	REGULAR	
	TÁXI AÉREO	
aeronave_voo_origem	designativo OACI=631	Categórica nominal
aeronave_voo_destino	designativo OACI=643	Categórica nominal
aeronave_fase_operacao	diversas fase de operacao=32	Categórica nominal

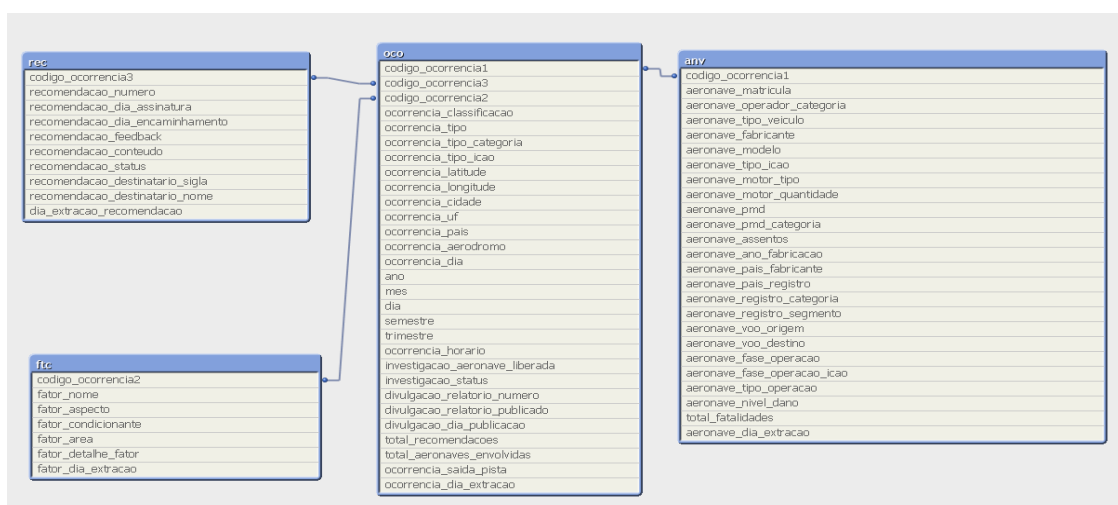
aeronave_fase_operacao_icao	diversas fase de operacao=32	Categórica nominal
aeronave_tipo_operacao	INDETERMINADA	Categórica nominal
	OPERAÇÃO AGRÍCOLA	
	OPERAÇÃO ESPECIALIZADA	
	OPERAÇÃO POLICIAL	
	TÁXI AÉREO	
	VOO DE INSTRUÇÃO	
	VOO EXPERIMENTAL	
	VOO NÃO REGULAR	
	VOO PRIVADO	
	VOO REGULAR	
aeronave_nivel_dano	DESTRUÍDA	Categórica nominal
	INDETERMINADO	
	LEVE	
	NENHUM	
	SUBSTANCIAL	
total_fatalidades	numero	numerico
aeronave_dia_extracao	data	data

3. Processamento/Tratamento de Dados

O trabalho será executado com as tabelas ocorrência e aeronave.

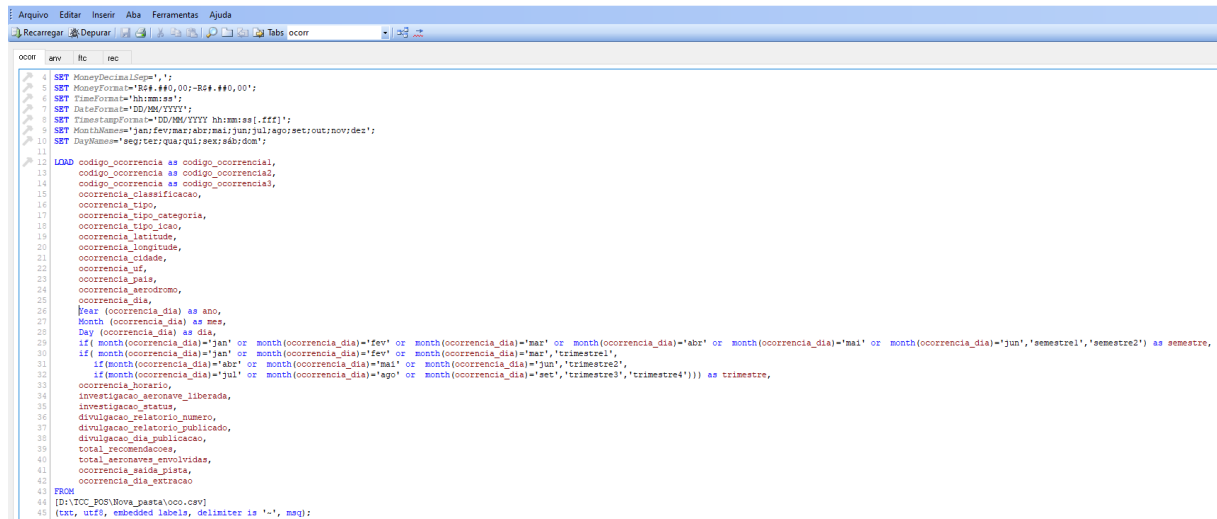
Após o relacionamento das tabelas no QlikView, através do campo código_ocorrendia, foram criadas novas variáveis na aba oco (figura 2): ano, mês, dia, semestre e trimestre. Estas novas variáveis podem fazer parte do KDD. Observar figuras abaixo.

Figura 1



Na figura 2, tem-se o script para criar as novas variáveis.

Figura 2



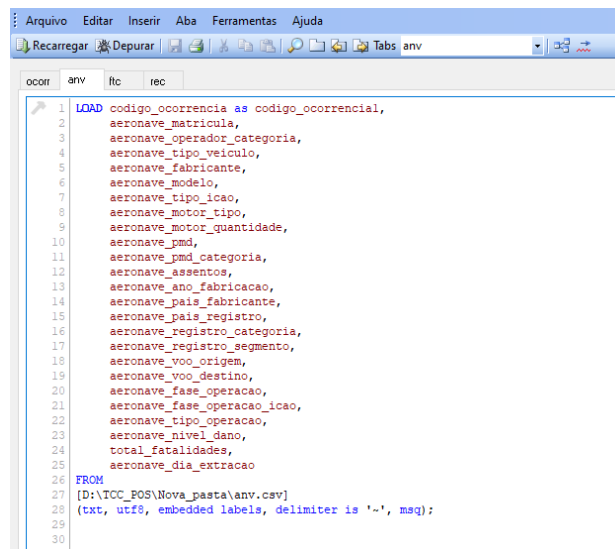
```

4 SET MoneyDecimalSep='.';
5 SET MoneyFormat='BRL,##0,00-;BRL,##0,00';
6 SET TimeFormat='hh:mm:ss';
7 SET DateFormat='DD/MM/YYYY';
8 SET TimestampFormat='DD/MM/YYYY hh:mm:ss[.fff]';
9 SET MonthNames='jan;fev;mar;abr;mai;jun;jul;ago;set;out;nov;dez';
10 SET DayNames='seg;ter;qua;qui;sex;sáb;dom';
11
12 LOAD
13   codigo_ocorrendia as codigo_ocorrendia,
14   codigo_ocorrendia as codigo_ocorrendia,
15   codigo_ocorrendia as codigo_ocorrendia,
16   ocorrendia_classificacao,
17   ocorrendia_tipo,
18   ocorrendia_tipo_categoria,
19   ocorrendia_tipo_icao,
20   ocorrendia_latitude,
21   ocorrendia_longitude,
22   ocorrendia_cidade,
23   ocorrendia_uf,
24   ocorrendia_pais,
25   ocorrendia_aerodromo,
26   ocorrendia_dia,
27   Year (ocorrendia_dia) as ano,
28   Month (ocorrendia_dia) as mes,
29   Day (ocorrendia_dia) as dia,
30   if (month(ocorrendia_dia)='jan' or month(ocorrendia_dia)='fev' or month(ocorrendia_dia)='mar' or month(ocorrendia_dia)='abr' or month(ocorrendia_dia)='mai' or month(ocorrendia_dia)='jun', 'semestre1', 'semestre2') as semestre,
31   if (month(ocorrendia_dia)='jan' or month(ocorrendia_dia)='fev' or month(ocorrendia_dia)='mar', 'trimestre1',
32   if (month(ocorrendia_dia)='abr' or month(ocorrendia_dia)='mai' or month(ocorrendia_dia)='jun', 'trimestre2',
33   if (month(ocorrendia_dia)='jul' or month(ocorrendia_dia)='ago' or month(ocorrendia_dia)='set', 'trimestre3', 'trimestre4')) as trimestre,
34   ocorrendia_horario,
35   investigacao_aeronave_liberada,
36   investigacao_status,
37   divulgacao_relatorio_numero,
38   divulgacao_relatorio_publicado,
39   divulgacao_dia_publicacao,
40   total_recomendacoes,
41   total_aeronaves_envolvidas,
42   ocorrendia_saida_pista,
43   ocorrendia_dia_extracao
44 FROM
45 [D:\TCC_POS\Nova_pasta\oco.csv]
46 (txt, utf8, embedded labels, delimiter is ',', msq);

```

A tabela aeronave permanece sem alteração.

Figura 3



```

1 LOAD
2   aeronave_matricula,
3   aeronave_operador_categoria,
4   aeronave_tipo_veiculo,
5   aeronave_fabricante,
6   aeronave_modelo,
7   aeronave_tipo_icao,
8   aeronave_motor_tipo,
9   aeronave_motor_quantidade,
10  aeronave_pmd,
11  aeronave_pmd_categoria,
12  aeronave_assentos,
13  aeronave_ano_fabricacao,
14  aeronave_pais_fabricante,
15  aeronave_pais_registro,
16  aeronave_registro_categoria,
17  aeronave_registro_segmento,
18  aeronave_voo_origem,
19  aeronave_voo_destino,
20  aeronave_fase_operacao,
21  aeronave_fase_operacao_icao,
22  aeronave_tipo_operacao,
23  aeronave_nivel_dano,
24  total_fatalidades,
25  aeronave_dia_extracao
26 FROM
27 [D:\TCC_POS\Nova_pasta\anv.csv]
28 (txt, utf8, embedded labels, delimiter is ',', msq);
29
30

```

No QlikView, foi criada uma tabela única através do menu Layout, Novo Objeto de Pasta e Tabela (figuras 4 e 5). Após foram inseridas as variáveis da tabela oco e tabela anv para forma um novo dataset com 5523 instâncias e 51 atributos. Observar as figuras 6 e 7.

Figura 4

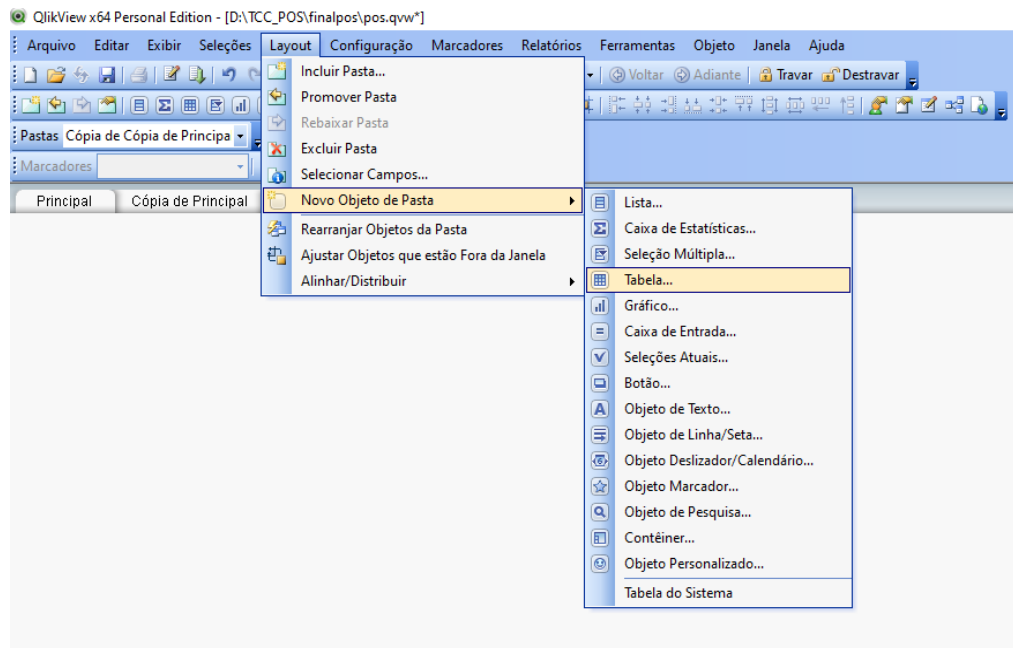


Figura 5

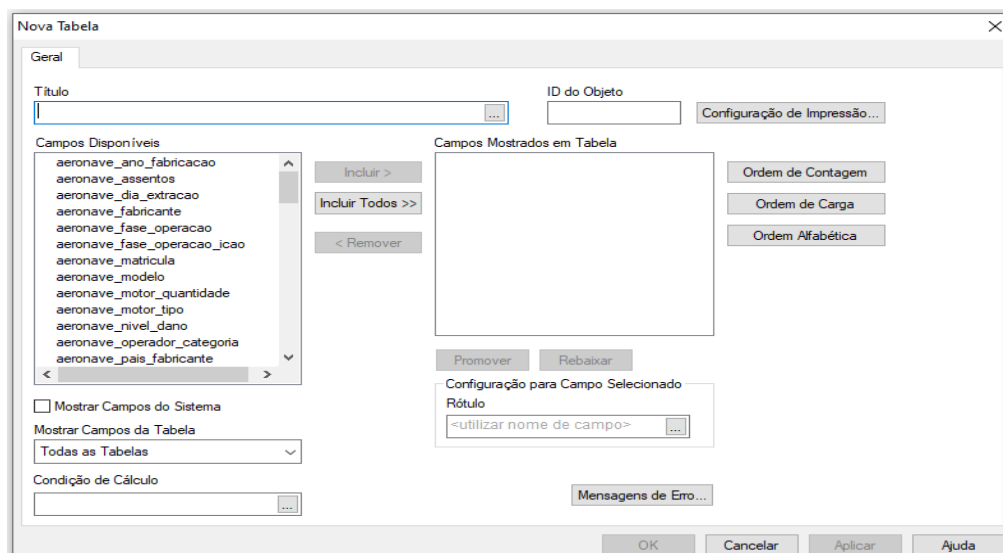


Figura 6

Propriedades da Tabela []

Geral | Classificar | Apresentação | Estilo | Número | Fonte | Layout | Título

Título: _____ ID do Objeto: TB03 Configuração de Impressão...

Campos Disponíveis:

- codigo_ocorrencia2
- codigo_ocorrencia3
- dia_extracao_recomendacao
- fator_area
- fator_aspecto
- fator_condicionante
- fator_detalhe_fator
- fator_dia_extracao
- fator_nome
- recomendacao_conteudo
- recomendacao_destinatario_nome
- recomendacao_destinatario_sigla
- recomendacao_dia_assinatura

Incluir > Incluir Todos >> < Remover

Campos Mostrados em Tabela:

- codigo_ocorrencia1
- ano
- mes
- dia
- semestre
- trimestre
- aeronave_motor_quantidade
- aeronave_tipo_operacao
- aeronave_motor_tipo
- aeronave_pmd_categoria
- aeronave_registro_segmento

Promover Rebaixar

Ordem de Contagem
Ordem de Carga
Ordem Alfabética

Configuração para Campo Selecionado

Rótulo: <utilizar nome de campo>

Mostrar Campos do Sistema ☐

Mostrar Campos da Tabela: Todas as Tabelas

Condição de Cálculo: _____ Mensagens de Erro...

OK Cancelar Aplicar Ajuda

Figura 7

[illegible]

Em uma base de dados, onde vários especialistas de segurança de voo inserem informações pertinentes as ocorrências aeronáuticas, observou-se dados ruidosos: incompletos e inconsistentes. E não foi possível detectar ruído de classe, que neste caso fica sob responsabilidade do órgão competente, CENIPA, para constatar algum ruído e fazer a devida correção na sua fonte de dados.

Na tabela 3, tem-se a lista de atributos excluídos por não ser relevante e por ter observações ruidosas, incompletos e inconsistentes. E os possíveis atributos para compor o modelo.

Dessa forma, os valores observados com triplo asterisco (***) nos atributos:

aeronave_motor_quantidade (36)

aeronave_motor_tipo (59)

aeronave_pmd_categoria (38)

aeronave_registro_segmento (83)

aeronave_registro_categoria (19)

foram substituídos por INDETERMINADA, de modo a configurar uma classe em um determinado atributo, e não perder informação caso fossem excluídas. Este procedimento agrega escolha dos atributos com relevância para construção do modelo preditivo.

A seguir a lista dos atributos, cor verde, para realizar o trabalho proposto:

Tabela 3

Excluídos	Possíveis atributos do modelo
codigo_ocorrencia1	
ano	
mes	
dia	
ocorrencia_aerodromo	
ocorrencia_cidade	
ocorrencia_horario	
ocorrencia_latITUDE	
ocorrencia_longitude	
ocorrencia_pais	semestre
ocorrencia_uf	trimestre
aeronave_assentos	aeronave_motor_quantidade
aeronave_fase_operacao	aeronave_tipo_operacao
aeronave_matricula	aeronave_motor_tipo
aeronave_modelo	aeronave_pmd_categoria
aeronave_pmd	aeronave_registro_segmento
aeronave_tipo_icao	ocorrencia_classificacao
ocorrencia_dia	ocorrencia_saida_pista
total_aeronaves_envolvidas	ocorrencia_tipo
total_fatalidades	ocorrencia_tipo_categoria
aeronave_ano_fabricacao	ocorrencia_tipo_icao
aeronave_dia_extracao	aeronave_fase_operacao_icao
aeronave_fabricante	aeronave_nivel_dano
aeronave_pais_fabricante	aeronave_operador_categoria
aeronave_pais_registro	aeronave_registro_categoria
aeronave_voo_destino	aeronave_tipo_veiculo
aeronave_voo_origem	
divulgacao_dia_publicacao	
divulgacao_relatorio_numero	
divulgacao_relatorio_publicado	
investigacao_aeronave_liberada	
investigacao_status	
ocorrencia_dia_extracao	
total_recomendacoes	

Agora o dataset possui 5352 instâncias e 17 atributos.

4. Análise e Exploração dos Dados

Por meio do QlikView, foram gerados tabelas e gráficos para exploração e análise de dados. Observa-se que, no atributo `aeronave_tipo_veiculo`, avião tem mais ocorrências aeronáuticas do que as demais, com 82%. As aeronaves bimotoras e monomotor, tem respectivamente 50% e 46,9%, registros de ocorrências aeronáuticas. E o total de incidente é maior do que acidente e incidente grave juntos, 56,2% e 43,8 %, respectivamente, ver figura 8. As aeronaves registradas como segmento particular e regular são as que se destacam, 29,7% e 21,1% respectivamente, envolvidas nas ocorrências. E os voos privados e voos regulares, que demonstra o tipo de operação em voo, representam 31,7% e 21,6% respectivamente. Percebe-se que em `aeronave_motor_quantidade` e `aeronave_registro_segmento` que as ocorrências com segmento particular e instrução estão concentradas em 30% com aeronaves monomotor. E as ocorrências com segmento particular e regular concentram-se em 33% com aeronaves bimotores (figura 12).

Figura 8

aeronave_tipo_veiculo <ul style="list-style-type: none"> AVIÃO 82.0% HELICÓPTERO 11.6% ULTRALEVE 5.3% INDETERMINADA 0.4% PLANADOR 0.3% ANFÍBIO 0.3% TRIKE 0.1% DIRIGÍVEL 0.0% HIDROAVIÃO 0.0% BALÃO 0.0% 	aeronave_motor_quantidade <ul style="list-style-type: none"> MONOMOTOR 50.0% BIMOTOR 46.9% TRIMOTOR 1.4% SEM TRAÇÃO 0.9% INDETERMINADA 0.7% QUADRIMOTOR 0.2%
aeronave_registro_segmento <ul style="list-style-type: none"> PARTICULAR 29.7% REGULAR 21.1% INSTRUÇÃO 14.5% TÁXI AÉREO 14.2% EXPERIMENTAL 7.5% AGRÍCOLA 5.4% ADMINISTRAÇÃO DIRETA 3.1% ESPECIALIZADA 2.0% INDETERMINADA 1.5% MÚLTIPLA 0.5% NÃO REGULAR 0.4% ADMINISTRAÇÃO INDIRETA 0.2% HISTÓRICA 0.0% 	aeronave_motor_tipo <ul style="list-style-type: none"> PISTÃO 56.0% JATO 21.2% TURBOÉLICE 12.0% TURBOEIXO 9.4% INDETERMINADA 1.1% SEM TRAÇÃO 0.3%
aeronave_nivel_dano <ul style="list-style-type: none"> NENHUM 34.6% LEVE 31.0% SUBSTANCIAL 27.3% DESTRUIDA 6.2% INDETERMINADO 0.8% 	aeronave_tipo_operacao <ul style="list-style-type: none"> VOO PRIVADO 31.7% VOO REGULAR 21.6% TÁXI AÉREO 14.0% VOO DE INSTRUÇÃO 14.0% OPERAÇÃO AGRÍCOLA 7.4% VOO EXPERIMENTAL 5.1% OPERAÇÃO POLICIAL 2.6% OPERAÇÃO ESPECIALIZADA 1.5% VOO NÃO REGULAR 1.1% INDETERMINADA 1.0%
aeronave_pmd_categoria <ul style="list-style-type: none"> ATÉ 2250 KG 54.9% ACIMA DE 5700 KG 27.0% ENTRE 2250 E 5700 KG 17.4% INDETERMINADA 0.7% 	ocorrencia_classificacao <ul style="list-style-type: none"> INCIDENTE 56.2% ACIDENTE 31.5% INCIDENTE GRAVE 12.3%
semestre <ul style="list-style-type: none"> semestre1 53.3% semestre2 46.7% 	ocorrencia_saida_pista <ul style="list-style-type: none"> NÃO 91.8% SIM 8.2%
trimestre <ul style="list-style-type: none"> trimestre1 28.2% trimestre2 25.1% trimestre3 23.7% trimestre4 23.0% 	aeronave_operador_categoria <ul style="list-style-type: none"> OPERADOR DE AERONAVE 44.0% OPERADOR PARTICULAR 38.3% AEROCLUBE 9.5% ESCOLA DE AVIAÇÃO 4.5% ÓRGÃO PÚBLICO 3.5% FABRICANTE AERONÁUT... 0.2%

Figura 9

ocorrencia_tipo	
FALHA DO MOTOR EM VOO	11.8%
ESTOURO DE PNEU	10.8%
FALHA OU MAU FUNCIONAMENTO DE SISTEMA / COMPONENTE	10.2%
PERDA DE CONTROLE NO SOLO	8.4%
PERDA DE CONTROLE EM VOO	6.3%
OUTROS	6.0%
COLISÃO COM AVE	5.8%
COM TREM DE POUSO	5.7%
COLISÃO COM OBSTÁCULO DURANTE A DECOLAGEM E POUSO	3.1%
TRÁFEGO AÉREO	2.7%
COLISÃO COM OBSTÁCULOS NO SOLO	2.1%
CAUSADO POR FENÔMENO METEOROLÓGICO EM VOO	2.0%
PANE SECA	1.8%
POUSO EM LOCAL NÃO PREVISTO	1.8%
INDETERMINADO	1.8%
POUSO SEM TREM	1.6%
POUSO BRUSCO	1.5%
COM PARA-BRISAS / JANELA / PORTA	1.2%
PERDA DE COMPONENTE EM VOO	1.1%

ocorrencia_tipo_categoria	
FALHA OU MAU FUNCIONAMENTO DE SISTEMA / COMPONENTE	31.2%
FALHA OU MAU FUNCIONAMENTO DO MOTOR	13.6%
OUTROS	12.8%
PERDA DE CONTROLE NO SOLO	8.4%
PERDA DE CONTROLE EM VOO	6.3%
COLISÃO COM AVE	5.8%
CONTATO ANORMAL COM A PISTA	3.3%
PERDA DE SEPARAÇÃO / COLISÃO EM VOO	3.2%
COLISÃO COM OBSTÁCULO DURANTE A DECOLAGEM E POUSO	3.1%
COLISÃO NO SOLO	3.0%
COMBUSTÍVEL	2.0%
INDETERMINADO	1.8%
EXCURSÃO DE PISTA	1.7%
FOGO/FUMAÇA (SEM IMPACTO)	1.0%
OPERAÇÃO A BAIXA ALTITUDE	1.0%
VOO CONTROLADO CONTRA O TERRENO	0.5%
POUSO AQUÉM/ALÉM DA PISTA	0.3%
MÉDICO	0.3%
INCURSÃO EM PISTA	0.3%
COLISÃO COM FAUNA	0.3%
OPERAÇÕES NO SOLO	0.1%
AERÓDROMO	0.1%
TURBULÊNCIA	0.1%

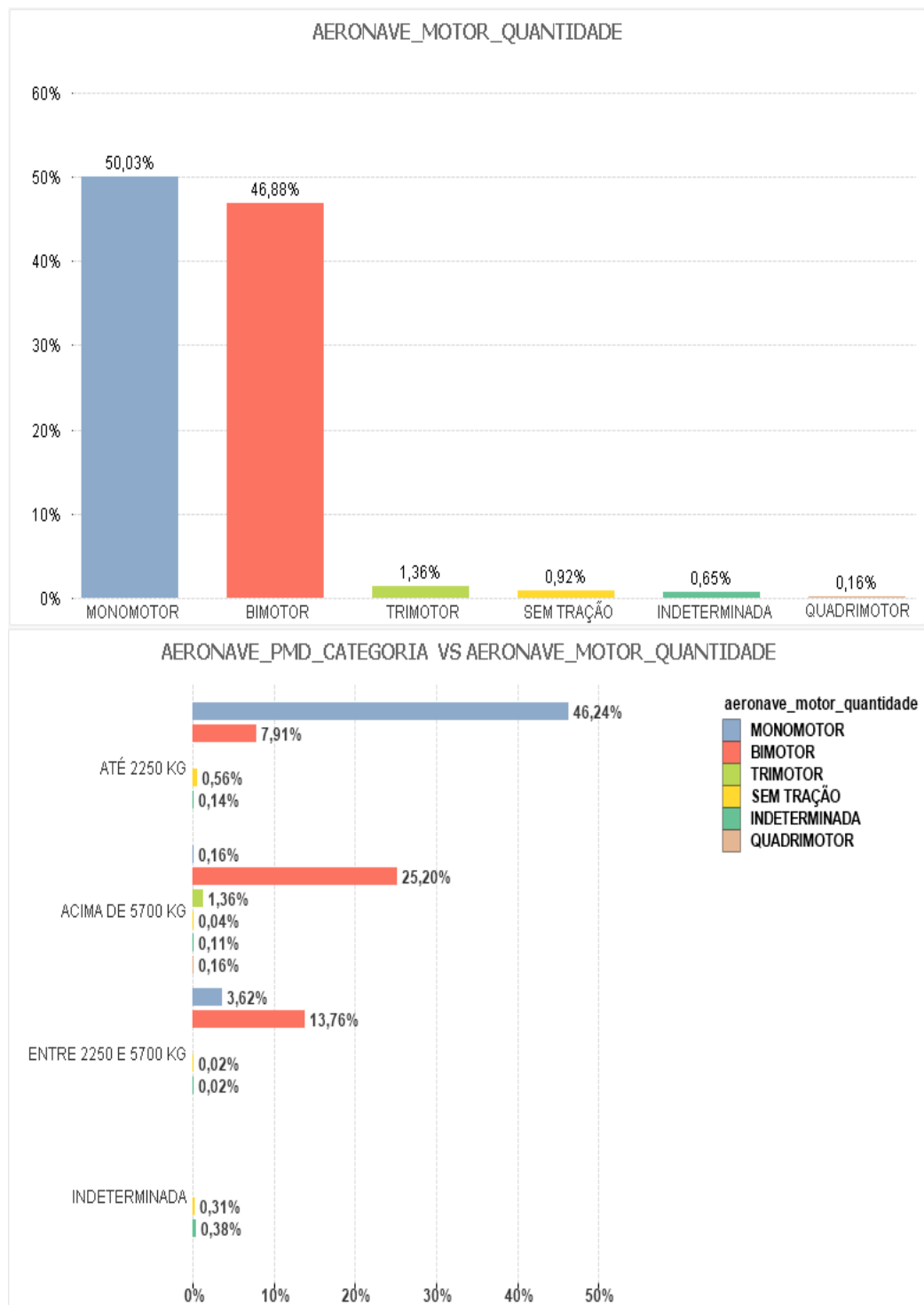
aeronave_fase_operacao_icao	
LANDING	27.8%
EN ROUTE	24.1%
TAKEOFF	16.1%
APPROACH	9.5%
TAXI	7.8%
MANEUVERING	7.5%
UNKNOWN	3.9%
STANDING	3.0%
PUSHBACK/TOWING	0.1%
UNCONTROLLED DESCENT	0.1%
EMERGENCY DESCENT	0.0%
NOT APLICABLE	0.0%

aeronave_registro_categoria	
TPP	29.7%
TPR	21.1%
PRI	14.2%
TPX	14.2%
PET	7.3%
SAE-AG	5.4%
ADE	2.4%
SAE	1.5%
EXT	1.2%
ADF	0.6%

aeronave_motor_tipo	
PISTÃO	56.0%
JATO	21.2%
TURBOÉLICE	12.0%
TURBOEIXO	9.4%
INDETERMINADA	1.1%
SEM TRAÇÃO	0.3%

ocorrencia_tipo_icao	
SCF-NP	31.2%
SCF-PP	13.6%
OTHR	12.8%
LOC-G	8.4%
LOC-I	6.3%
BIRD	5.8%
ARC	3.3%
MAC	3.2%
CTOL	3.1%
GCOL	3.0%
FUEL	2.0%

Figura 10



As ocorrências aeronáuticas com aeronaves monomotor (46,2%) são da categoria de peso máximo de decolagem até 2250KG, indicando pertencer ao segmento particular e instrução (aviação geral) – figura 10.

Figura 11

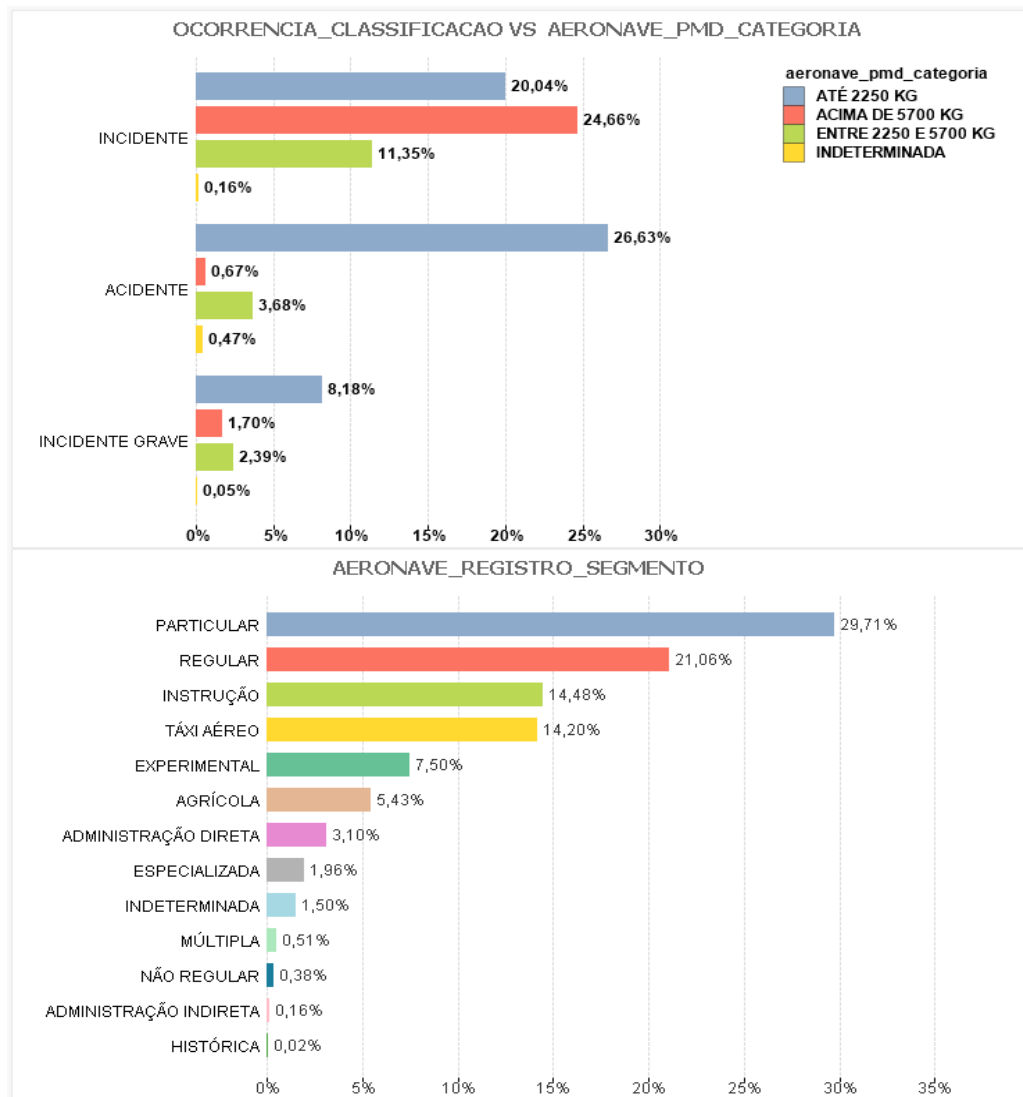
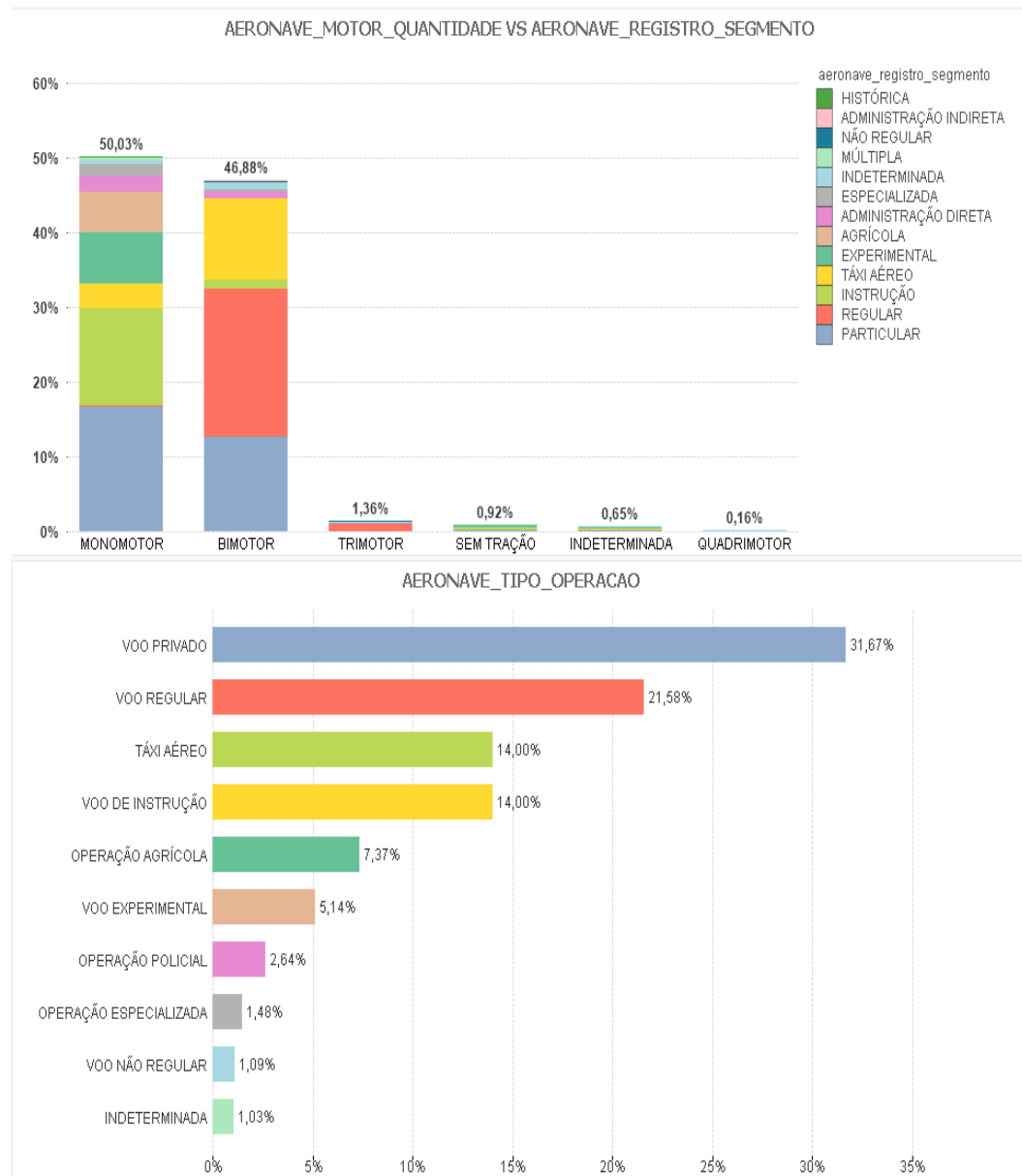


Figura 12



Conforme o RAB, 65% e 26% respectivamente da frota no Brasil, é composta pela aviação geral e aviação experimental. Dessa forma, o estudo proposto para

elaboração de um modelo preditivo é baseado no quantitativo de motor da aeronave como variável alvo. Visto que, as ocorrências registradas para aeronaves **monomotor** e **bimotor** estão bem equilibradas, não havendo preocupação com desbalanceamento que podem causar problemas em um modelo de classificação. Com isso, o dataset possui 5352 instâncias, 17 atributos e a variável alvo (aeronave_motor_quantidade) com duas classes, isto significa que aeronaves trimotor, sem tração, indeterminada e quadrimotor não foram parte do modelo. Então as porcentagens de monomotor e bimotor, que eram de 50% e 46,9%, agora estão em 51,6% e 48,4% respectivamente.

5. Criação de Modelos de Machine Learning

O Orange Versão: 3.24.1 é em código aberto, escrito em Python, Cython, C++ e C, que vai ser utilizado para produção de modelos preditivos de classificação a partir de uma base de dados históricos, para depois prever o comportamento futuro através dos registros passados. O Orange possui uma interface visual, Orange Canvas, para interligar objetos, manipular dados e desenvolver modelos de classificação: Árvore de decisão, Floresta aleatória, Adaboost e Naive bayes.

Na interface gráfica inicial no Orange, a função **file** faz a leitura de um arquivo de entrada em Excel (.xlsx) e seleção dos atributos. O **Rank** fornece pontuação para as features para compor o modelo. O **Preprocess** permite transformar os valores de atributos e a **Tabela de dados** é para visualizar os valores transformados. Ver figura 13.

Figura 13

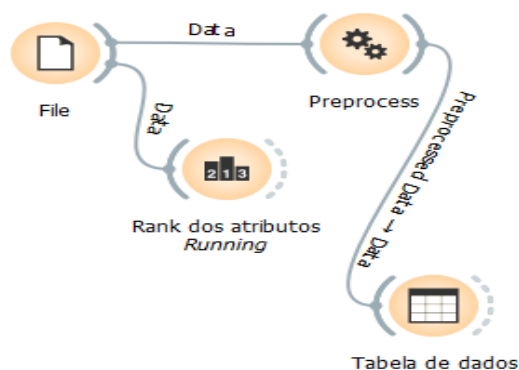


Figura 15

Data Table (1)

Info	aeronave_motor_quantidade	semestre=semestre1	semestre=semestre2
5352 instances (no missing values)	1 MONOMOTOR	1	0
233 features (no missing values)	2 BIMOTOR	1	0
Discrete class with 2 values (no missing values)	3 BIMOTOR	1	0
No meta attributes	4 BIMOTOR	1	0
Variables	5 MONOMOTOR	1	0
<input checked="" type="checkbox"/> Show variable labels (if present)	6 BIMOTOR	1	0
<input checked="" type="checkbox"/> Visualize numeric values	7 BIMOTOR	1	0
<input type="checkbox"/> Color by instance classes	8 BIMOTOR	1	0
Selection	9 MONOMOTOR	1	0
<input checked="" type="checkbox"/> Select full rows	10 MONOMOTOR	1	0
	11 BIMOTOR	1	0
	12 BIMOTOR	1	0
	13 BIMOTOR	1	0

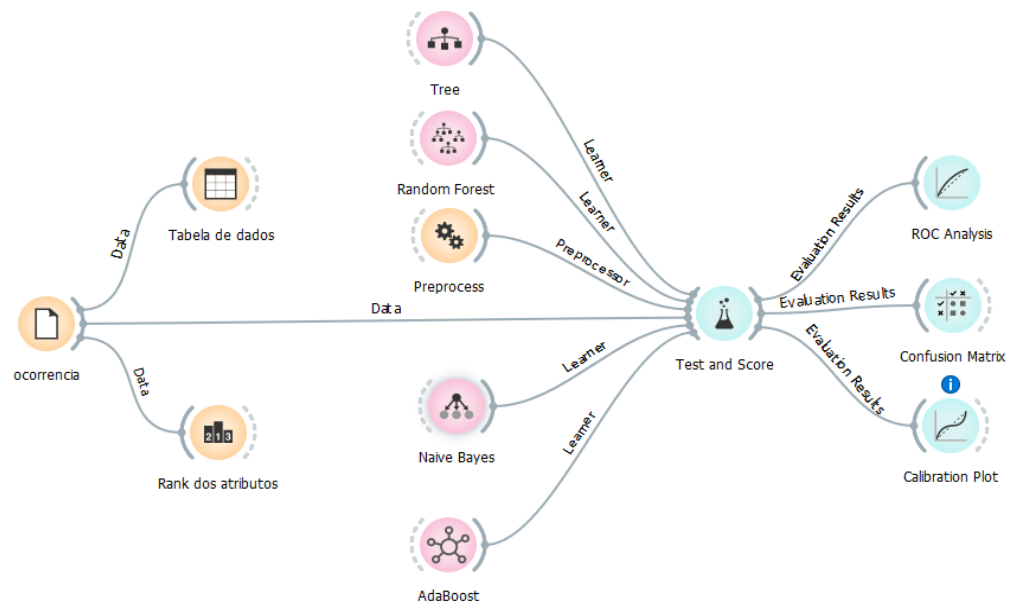
Figura 16

Data Table

Info	ave_motor_quant	semestre
5352 instances (no missing values)	1 MONOMOTOR	semestre1
16 features (no missing values)	2 BIMOTOR	semestre1
Discrete class with 2 values (no missing values)	3 BIMOTOR	semestre1
No meta attributes	4 BIMOTOR	semestre1
Variables	5 MONOMOTOR	semestre1
<input checked="" type="checkbox"/> Show variable labels (if present)	6 BIMOTOR	semestre1
<input type="checkbox"/> Visualize numeric values	7 BIMOTOR	semestre1
<input checked="" type="checkbox"/> Color by instance classes	8 BIMOTOR	semestre1
Selection	9 MONOMOTOR	semestre1
<input checked="" type="checkbox"/> Select full rows	10 MONOMOTOR	semestre1
	11 BIMOTOR	semestre1
	12 BIMOTOR	semestre1
	13 BIMOTOR	semestre1

Os seguintes métodos de classificadores serão utilizados para compor o modelo preditivo: árvore de decisão, floresta aleatória, adaboost e naive bayes. A interface no Orange (figura 17) mostra o esquema de ligações entre os classificadores e test and score (treinamento e teste de aprendizado de dados) com seus respectivos resultados de forma visual em gráficos das curvas ROC (ROC Analysis), matriz de confusão (Confusion Matrix) e gráfico de calibração (Calibration Plot).

Figura 17



Após clicar em Rank dos atributos, figura 17, é aberto uma janela. Tem-se os métodos de filtragem de atributos para compor o modelo de classificação: Info.gain, Gain ratio, Gini e ReliefF.

- Info.gain -mede a semelhança entre o atributo de uma classe, com todas as outras classes, diminuindo o grau de incerteza (ANJOS, 2019).
- Gain ratio – seleciona atributos com maior possibilidades de valores (NETTO, 2013).
- ReliefF – mede a relevância dos atributos em relação à classe usando uma medida de distância para ordenar os atributos (Huei Diana Lee, 2003).
- Gini – é a desigualdade entre valores de uma distribuição de frequência (Orange Widget catalog).
- Qui-quadrado- teste de associação aplicado para selecionar as features do modelo.

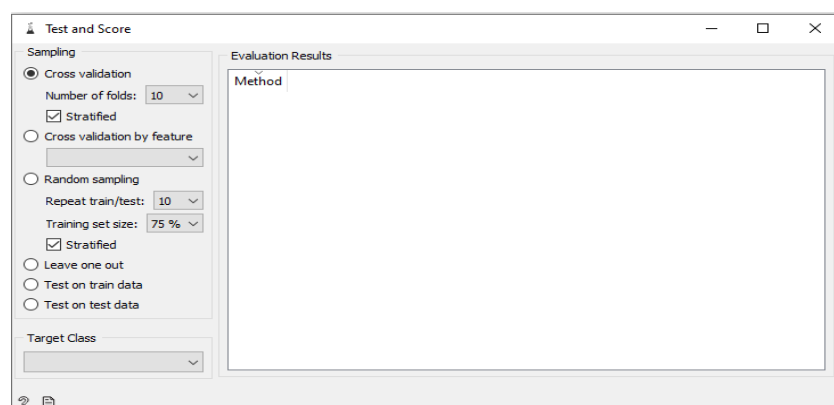
Figura 18

Rank

Scoring Methods		#	Info. gain	Gain ratio	Gini	χ^2	ReliefF
<input checked="" type="checkbox"/> Information Gain	<input checked="" type="checkbox"/> aeronave_pmd_categoria	3	0.517	0.364	0.298	142.919	0.390
<input checked="" type="checkbox"/> Information Gain Ratio	<input checked="" type="checkbox"/> aeronave_registro_categoria	33	0.461	0.163	0.250	2641.371	0.300
<input checked="" type="checkbox"/> Gini Decrease	<input checked="" type="checkbox"/> aeronave_registro_segmento	13	0.453	0.166	0.246	1933.280	0.294
<input type="checkbox"/> ANOVA	<input checked="" type="checkbox"/> aeronave_tipo_operacao	10	0.442	0.167	0.241	736.887	0.292
<input checked="" type="checkbox"/> χ^2	<input checked="" type="checkbox"/> aeronave_motor_tipo	4	0.363	0.224	0.207	14.256	0.278
<input checked="" type="checkbox"/> ReliefF	<input checked="" type="checkbox"/> aeronave_operador_categoria	6	0.207	0.117	0.131	44.890	0.098
<input type="checkbox"/> FCBF	<input checked="" type="checkbox"/> ocorrencia_tipo	71	0.193	0.042	0.117	897.361	0.146
	<input checked="" type="checkbox"/> ocorrencia_classificacao	3	0.182	0.133	0.117	238.185	0.002
	<input checked="" type="checkbox"/> ocorrencia_tipo_categoria	29	0.152	0.046	0.094	504.162	0.110
	<input checked="" type="checkbox"/> ocorrencia_tipo_icao	29	0.152	0.046	0.094	38.263	0.110
	<input checked="" type="checkbox"/> aeronave_nivel_dano	5	0.148	0.079	0.096	21.609	0.020
	<input checked="" type="checkbox"/> aeronave_fase_operacao_icao	12	0.073	0.027	0.043	26.257	0.114
	<input checked="" type="checkbox"/> aeronave_tipo_veiculo	7	0.057	0.067	0.029	258.019	0.114
	<input checked="" type="checkbox"/> ocorrencia_saida_pista	2	0.005	0.012	0.003	33.964	0.028
	<input checked="" type="checkbox"/> trimestre	4	0.000	0.000	0.000	0.314	0.038
	<input checked="" type="checkbox"/> semestre	2	0.000	0.000	0.000	0.075	0.038

Os métodos de filtragem, figura 18, foram testados. E isto acontece no test and score que faz a avaliação da qualidade dos classificadores, e suporta vários métodos de amostragem: Cross-validation (Validação Cruzada) e random sampling (split-sample ou divisão simples), esta última divide os dados por amostragem de treino e teste, neste caso (75% e 25%), que pode ser repetido por um número específico de vezes, para este trabalho são 10. O Cross-validation é usado para evitar sobreajuste (modelo que se ajusta muito bem aos dados observados, porém não é eficaz para prever novos dados), sendo usado para esse processo 10 folds. A amostragem estratificada é usada tanto no random sample quanto no cross validation, de modo particionar o conjunto de dados na mesma proporção entre as classes existentes dos dados original. Clicando ou abrindo o Test and Score, figura 19, tem-se a interface no Orange.

Figura 19



Dessa forma, as features selecionadas para compor o modelo e a tabela composta com as variáveis dummy podem ser observadas na figura 20 e 21.

Figura 20

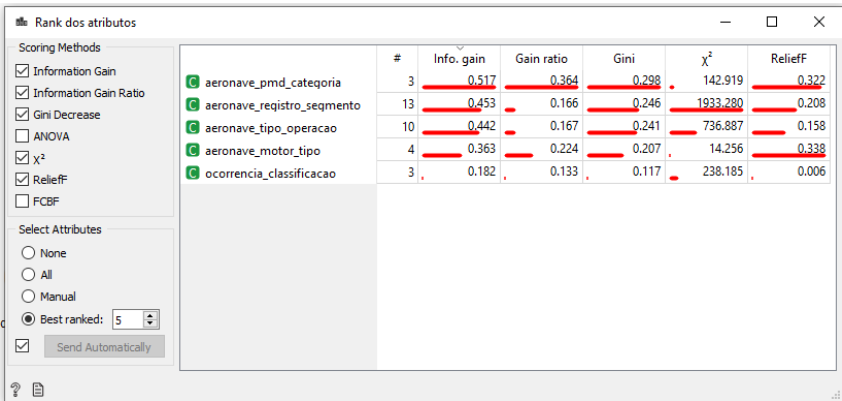
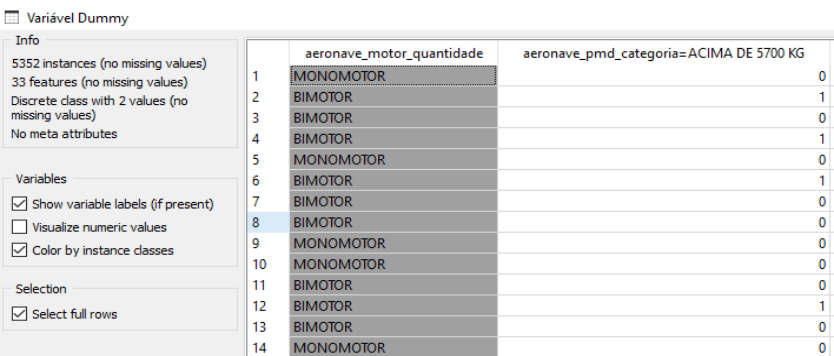


Figura 21



Após a seleção das features com maiores pontuações para cada método de seleção, tem-se os valores observados no test and score: AUC (Área sob a curva ROC), CA (Acurácia), F1 (F-score), Precision (Precisão) e Recall para os classificadores árvore de decisão, floresta aleatória, naive bayes e AdaBoost, a fim de obter o melhor modelo. Observar figuras 22 e 23, random sampling e cross validation respectivamente.

Figura 22

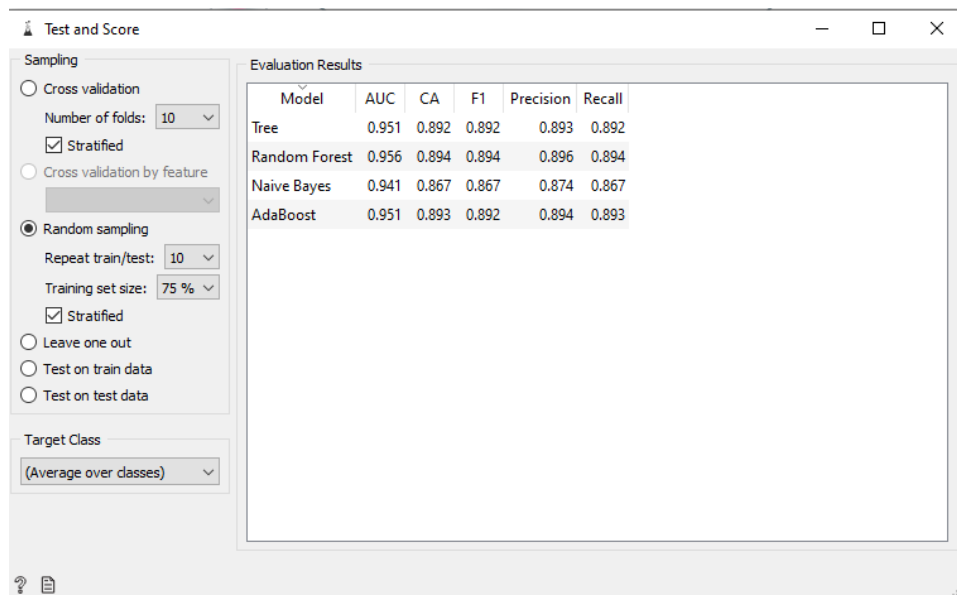
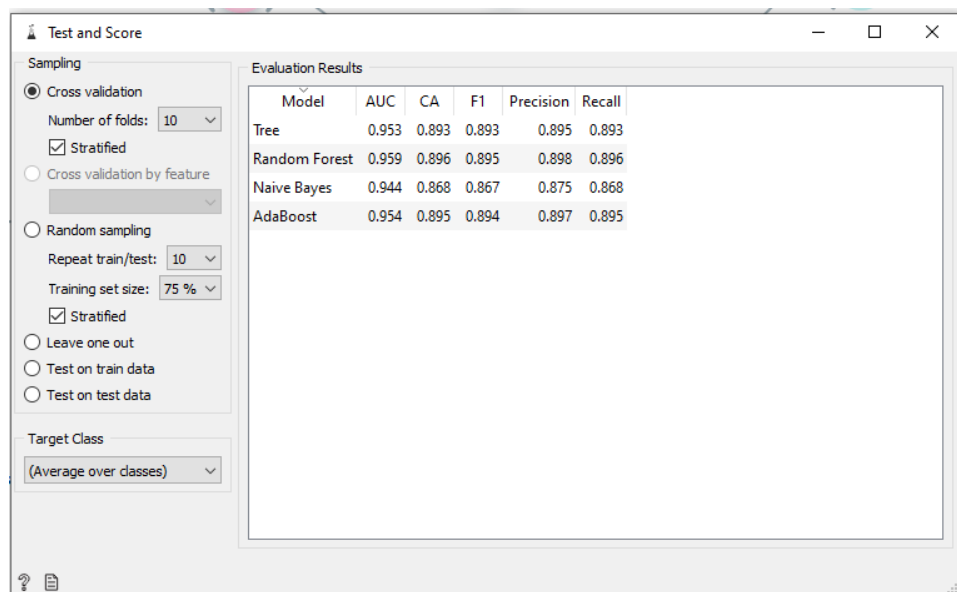


Figura 23



6. Apresentação dos Resultados

Dentre as duas metodologias de avaliação de classificadores: random sampling (divisão simples) e cross validation, optou-se por validação cruzada por ser

amplamente usada para validar a precisão dos modelos e comparar o desempenho de vários modelos (Análise Preditiva para leigos, 2019).

Os classificadores são as mais frequentes técnicas de aprendizado de máquina usados para problemas do mundo real, e a forma de compreender e avaliar o desempenho desses classificadores é através de uma matriz de confusão por meio de métricas:

CA (Acurácia) – percentual de predições corretas.

Recall (Revocação ou sensibilidade) – porcentagem de valores positivos classificados corretamente.

Precision (Precisão ou preditividade positiva) – porcentagem de predições positivas que estavam corretas.

F1 (F-score) – relação entre a precisão e revocação, isto é, média harmônica de revocação e precisão.

Logo abaixo, a tabela 4 para interpretação da matriz de confusão.

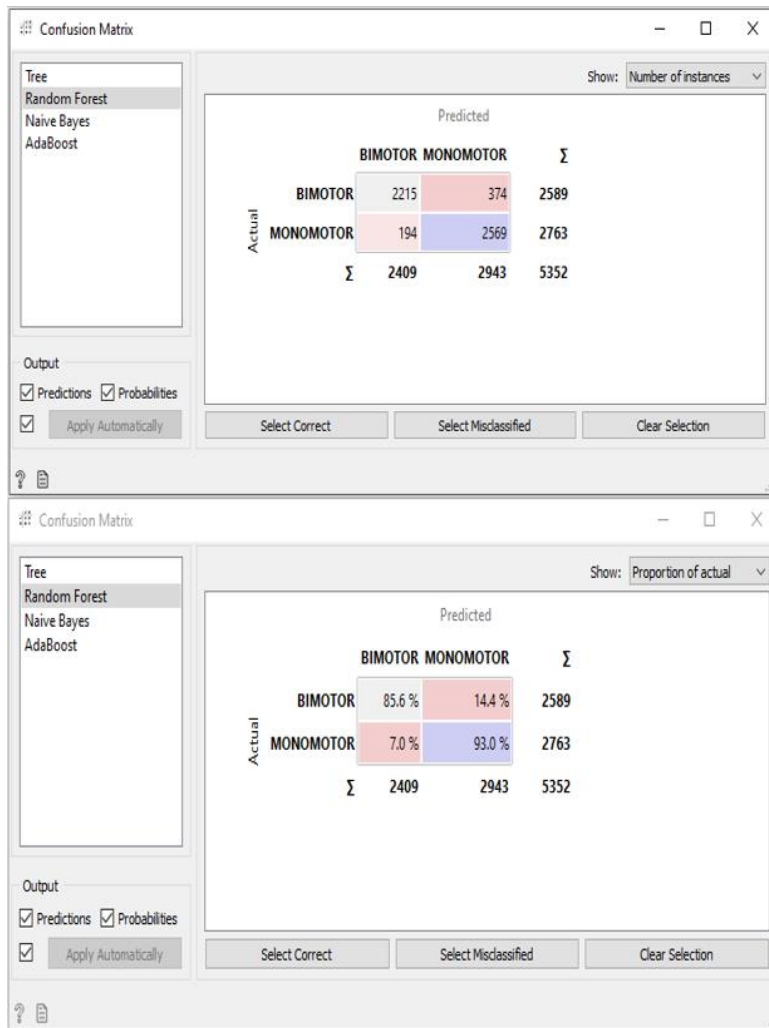
Tabela 4

	POSITIVO	NEGATIVO
POSITIVO	Verdadeiros positivos (VP)	Falso negativos (FN)
NEGATIVO	Falsos positivos (FP)	Verdadeiros negativos (VN)

Neste contexto, o modelo Random Forest foi escolhido por ter um acurácia de 89%, isto é, o algoritmo floresta aleatória obteve 89% das classificações corretas. Observando os valores das métricas, percebe-se quantitativos semelhantes tanto para divisão simples quanto validação cruzada (figuras 22 e 23), demonstrando equilíbrio entre as duas metodologias de avaliação de classificadores com os dados selecionados para este trabalho. E as métricas para o classificador Random Forest foram: F1=89,5%, Precisão=89,8%, Revocação=89,6% e AUC=95,9%.

Continuando com os resultados: ocorrências aeronáuticas com aeronaves bimotores foram classificadas corretamente com 85,6%, e 7% como falso positivo, isto é, 194 aeronaves bimotores foram classificadas como aeronave monomotor. O classificador acerta 93% das ocorrências com aeronaves monomotor. Observar figura 24.

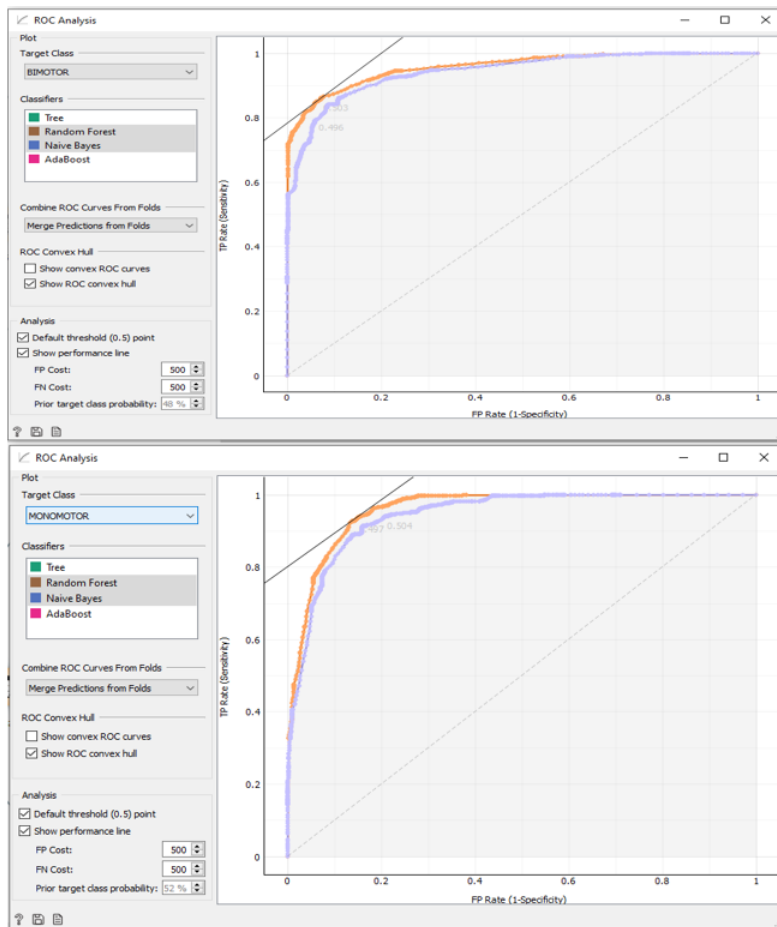
Figura 24



Nesta parte, far-se-á um comparativo entre os resultados da Curva ROC e Calibratin Plot para os classificadores Random Foreste e Naive bayes, sendo este último com a menor acurácia com 86%.

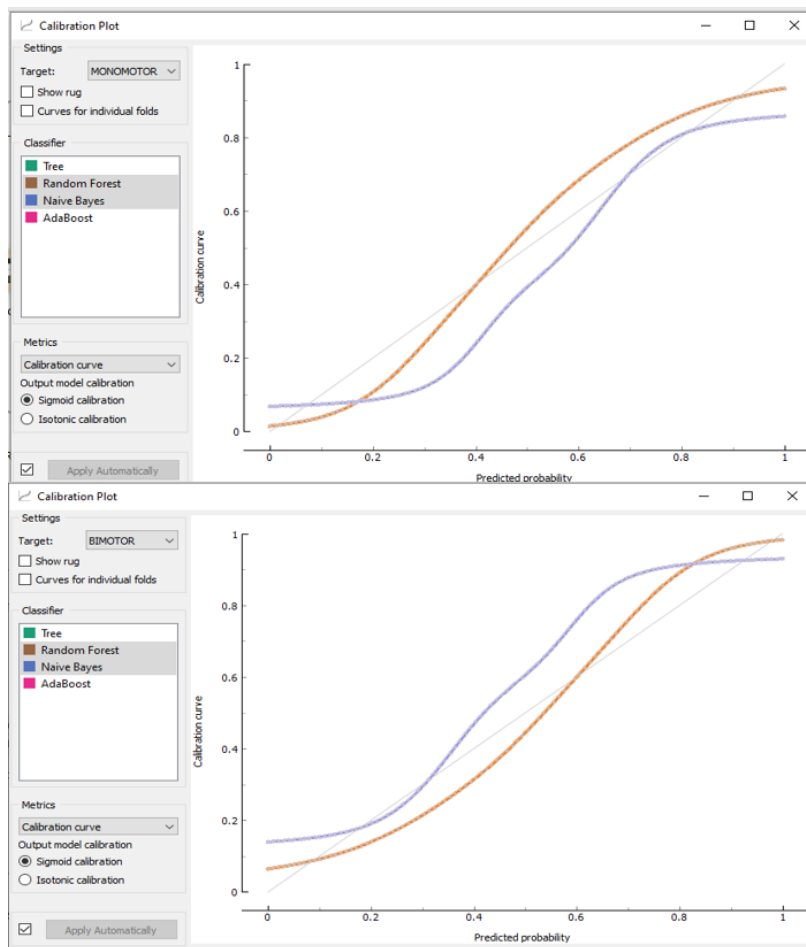
A Curva ROC será perfeita caso o tracejado chegue o mais próximo do canto superior esquerdo. E no caso de dois classificadores, o traçado que estiver mais à esquerda e superior de outro indica um desempenho melhor, isto é, tem uma maior taxa de verdadeiros positivos e uma menor taxa de falsos positivos. Neste caso, o classificador Random Forest com tracejado amarelo tem desempenho melhor para fazer previsões corretas do que o outro classificador comparado, tanto para variável alvo: monomotor quanto bimotor. Figura 25.

Figura 25



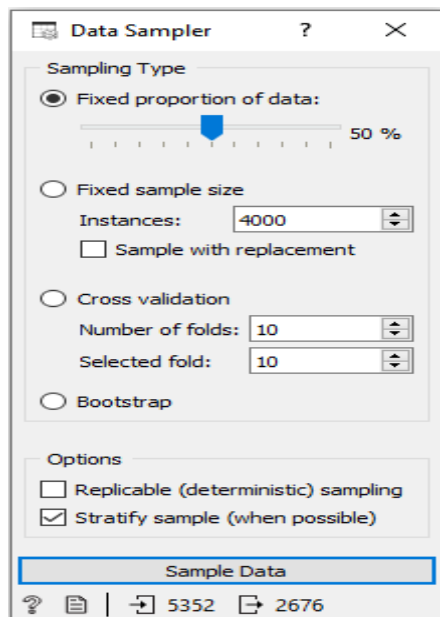
O gráfico de calibração (Calibration Plot) mostra a relação entre as previsões de probabilidade dos classificadores e as probabilidades reais da classe (Orange Widget catalog). Quanto mais próximo a curva do classificador se aproximar da reta diagonal, mais precisas são suas probabilidades de previsão. Observando a figura 26, é perceptível a curva de cor amarela mais próximo a linha diagonal, que corresponde ao classificador random foreste.

Figura 26



Já escolhido os dois classificadores, pode-se fazer a predição através da função Data Sampler. Os dados podem ser fixados proporcionalmente ou numericamente, e com ou sem reposição. Caso tenha que fixar 50% dos dados de ocorrências aleatoriamente com a opção estratificada, pode-se fazer predições com as repartições – Figura 27. Neste caso, ficou dividido em 2676 instâncias. A função preprocess continua com a codificação de indicadores para as variáveis dummy - figura 28.

Figura 27



Abaixo, a interface no Orange para fazer as previsões – Figura 28. Clicando duas vezes na ligação entre Data Sampler e Prediction, aparecerá duas opções conforme figura 29, isto decidirá os resultados de previsões: Data sample -> Data (previsões com a primeira partição dos dados) e Remaining data -> Data (previsões com os dados restantes). Logo após, as previsões nas tabelas 5 e 6.

Figura 28

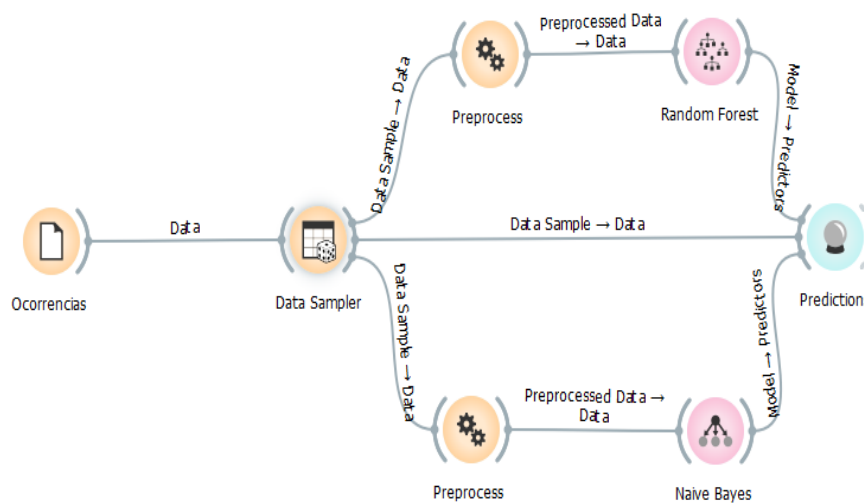


Figura 29

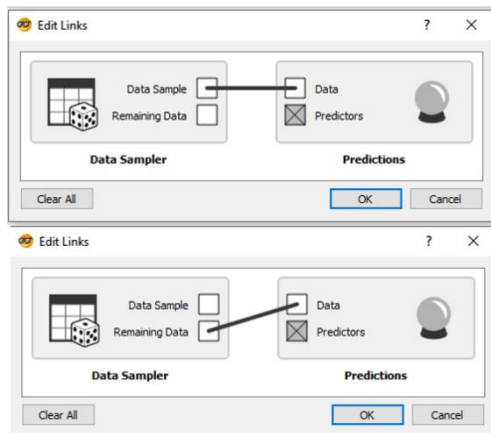


Tabela 5

	Naive Bayes	Random Forest	aeronave_motor_quantidade	aeronave_tipo_operacao	aeronave_motor_tipo	aeronave_pmd_categoria	aeronave_registro_segmento	ocorrencia_classificacao
1	1.00 : 0.00 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	VOO REGULAR	JATO	ACIMA DE 5700 KG	REGULAR	ACIDENTE
2	1.00 : 0.00 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	VOO REGULAR	TURBOEIXO	ACIMA DE 5700 KG	REGULAR	INCIDENTE
3	0.00 : 1.00 → MONOMOTOR	0.11 : 0.89 → MONOMOTOR	MONOMOTOR	VOO DE INSTRUÇÃO	PISTÃO	ATÉ 2250 KG	INSTRUÇÃO	INCIDENTE
4	0.46 : 0.54 → MONOMOTOR	0.76 : 0.24 → BIMOTOR	BIMOTOR	TÁXI AÉREO	PISTÃO	ATÉ 2250 KG	TÁXI AÉREO	INCIDENTE
5	0.00 : 1.00 → MONOMOTOR	0.06 : 0.94 → MONOMOTOR	MONOMOTOR	VOO DE INSTRUÇÃO	PISTÃO	ATÉ 2250 KG	INSTRUÇÃO	INCIDENTE GRAVE
6	0.00 : 1.00 → MONOMOTOR	0.29 : 0.71 → MONOMOTOR	BIMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	PARTICULAR	INCIDENTE GRAVE
7	0.02 : 0.98 → MONOMOTOR	0.33 : 0.67 → MONOMOTOR	MONOMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	PARTICULAR	INCIDENTE
8	1.00 : 0.00 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	VOO REGULAR	JATO	ACIMA DE 5700 KG	REGULAR	INCIDENTE
9	0.81 : 0.19 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	VOO PRIVADO	PISTÃO	ENTRE 2250 E 5700 KG	PARTICULAR	INCIDENTE
10	0.00 : 1.00 → MONOMOTOR	0.00 : 1.00 → MONOMOTOR	MONOMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	EXPERIMENTAL	ACIDENTE
11	0.00 : 1.00 → MONOMOTOR	0.07 : 0.93 → MONOMOTOR	BIMOTOR	VOO DE INSTRUÇÃO	PISTÃO	ATÉ 2250 KG	INSTRUÇÃO	ACIDENTE
12	0.00 : 1.00 → MONOMOTOR	0.15 : 0.85 → MONOMOTOR	MONOMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	PARTICULAR	ACIDENTE
13	1.00 : 0.00 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	VOO REGULAR	JATO	ACIMA DE 5700 KG	REGULAR	INCIDENTE
14	0.36 : 0.64 → MONOMOTOR	0.44 : 0.56 → MONOMOTOR	MONOMOTOR	OPERAÇÃO POLICIAL	TURBOEIXO	ENTRE 2250 E 5700 KG	ADMINISTRAÇÃO DIRETA	ACIDENTE
15	1.00 : 0.00 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	VOO REGULAR	JATO	ACIMA DE 5700 KG	REGULAR	INCIDENTE
16	1.00 : 0.00 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	VOO REGULAR	JATO	ACIMA DE 5700 KG	REGULAR	INCIDENTE
17	1.00 : 0.00 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	TÁXI AÉREO	TURBOEIXO	ACIMA DE 5700 KG	TÁXI AÉREO	INCIDENTE
18	0.00 : 1.00 → MONOMOTOR	0.15 : 0.85 → MONOMOTOR	MONOMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	PARTICULAR	ACIDENTE
19	0.02 : 0.98 → MONOMOTOR	0.33 : 0.67 → MONOMOTOR	MONOMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	PARTICULAR	INCIDENTE
20	0.02 : 0.98 → MONOMOTOR	0.33 : 0.67 → MONOMOTOR	MONOMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	PARTICULAR	INCIDENTE
21	1.00 : 0.00 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	VOO REGULAR	JATO	ACIMA DE 5700 KG	REGULAR	INCIDENTE
22	1.00 : 0.00 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	VOO REGULAR	JATO	ACIMA DE 5700 KG	INDETERMINADA	INCIDENTE
23	1.00 : 0.00 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	TÁXI AÉREO	PISTÃO	ENTRE 2250 E 5700 KG	TÁXI AÉREO	INCIDENTE
24	0.00 : 1.00 → MONOMOTOR	0.15 : 0.85 → MONOMOTOR	MONOMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	PARTICULAR	ACIDENTE
25	1.00 : 0.00 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	TÁXI AÉREO	TURBOEIXO	ACIMA DE 5700 KG	TÁXI AÉREO	INCIDENTE

Tabela 6

	Naive Bayes	Random Forest	aeronave_motor_quantidade	aeronave_tipo_operacao	aeronave_motor_tipo	aeronave_pmd_categoria	aeronave_registro_segmento	ocorrencia_classificacao
1	0.81 : 0.19 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	VOO PRIVADO	PISTÃO	ENTRE 2250 E 5700 KG	PARTICULAR	INCIDENTE
2	0.00 : 1.00 → MONOMOTOR	0.07 : 0.93 → MONOMOTOR	MONOMOTOR	VOO DE INSTRUÇÃO	PISTÃO	ATÉ 2250 KG	INSTRUÇÃO	ACIDENTE
3	0.00 : 1.00 → MONOMOTOR	0.07 : 0.93 → MONOMOTOR	MONOMOTOR	VOO DE INSTRUÇÃO	PISTÃO	ATÉ 2250 KG	INSTRUÇÃO	ACIDENTE
4	1.00 : 0.00 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	TÁXI AÉREO	JATO	ACIMA DE 5700 KG	TÁXI AÉREO	INCIDENTE
5	0.00 : 1.00 → MONOMOTOR	0.29 : 0.71 → MONOMOTOR	MONOMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	PARTICULAR	INCIDENTE GRAVE
6	1.00 : 0.00 → BIMOTOR	0.64 : 0.36 → BIMOTOR	BIMOTOR	TÁXI AÉREO	TURBOÉLICE	ENTRE 2250 E 5700 KG	TÁXI AÉREO	INCIDENTE
7	0.02 : 0.98 → MONOMOTOR	0.33 : 0.67 → MONOMOTOR	BIMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	PARTICULAR	INCIDENTE
8	0.00 : 1.00 → MONOMOTOR	0.07 : 0.93 → MONOMOTOR	MONOMOTOR	VOO DE INSTRUÇÃO	PISTÃO	ATÉ 2250 KG	INSTRUÇÃO	ACIDENTE
9	1.00 : 0.00 → BIMOTOR	0.64 : 0.36 → BIMOTOR	BIMOTOR	TÁXI AÉREO	TURBOÉLICE	ENTRE 2250 E 5700 KG	TÁXI AÉREO	INCIDENTE
10	0.00 : 1.00 → MONOMOTOR	0.11 : 0.89 → MONOMOTOR	MONOMOTOR	VOO DE INSTRUÇÃO	PISTÃO	ATÉ 2250 KG	INSTRUÇÃO	INCIDENTE
11	0.01 : 0.99 → MONOMOTOR	0.00 : 1.00 → MONOMOTOR	MONOMOTOR	VOO PRIVADO	TURBOEIXO	ATÉ 2250 KG	PARTICULAR	ACIDENTE
12	0.02 : 0.98 → MONOMOTOR	0.33 : 0.67 → MONOMOTOR	BIMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	PARTICULAR	INCIDENTE
13	0.46 : 0.54 → MONOMOTOR	0.76 : 0.24 → BIMOTOR	BIMOTOR	TÁXI AÉREO	PISTÃO	ATÉ 2250 KG	TÁXI AÉREO	INCIDENTE
14	1.00 : 0.00 → BIMOTOR	0.84 : 0.16 → BIMOTOR	MONOMOTOR	VOO PRIVADO	TURBOÉLICE	ENTRE 2250 E 5700 KG	PARTICULAR	INCIDENTE
15	1.00 : 0.00 → BIMOTOR	0.98 : 0.02 → BIMOTOR	BIMOTOR	TÁXI AÉREO	TURBOEIXO	ENTRE 2250 E 5700 KG	TÁXI AÉREO	INCIDENTE
16	0.00 : 1.00 → MONOMOTOR	0.15 : 0.85 → MONOMOTOR	MONOMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	PARTICULAR	ACIDENTE
17	0.00 : 1.00 → MONOMOTOR	0.15 : 0.85 → MONOMOTOR	MONOMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	PARTICULAR	ACIDENTE
18	0.02 : 0.98 → MONOMOTOR	0.33 : 0.67 → MONOMOTOR	MONOMOTOR	VOO PRIVADO	PISTÃO	ATÉ 2250 KG	PARTICULAR	INCIDENTE
19	0.00 : 1.00 → MONOMOTOR	0.00 : 1.00 → MONOMOTOR	MONOMOTOR	VOO EXPERIMENTAL	PISTÃO	ATÉ 2250 KG	EXPERIMENTAL	ACIDENTE
20	0.95 : 0.05 → BIMOTOR	0.68 : 0.32 → BIMOTOR	BIMOTOR	VOO PRIVADO	TURBOÉLICE	ENTRE 2250 E 5700 KG	PARTICULAR	INCIDENTE GRAVE
21	1.00 : 0.00 → BIMOTOR	1.00 : 0.00 → BIMOTOR	BIMOTOR	VOO REGULAR	JATO	ACIMA DE 5700 KG	REGULAR	INCIDENTE
22	0.35 : 0.65 → MONOMOTOR	0.00 : 1.00 → MONOMOTOR	MONOMOTOR	VOO PRIVADO	TURBOEIXO	ATÉ 2250 KG	PARTICULAR	INCIDENTE
23	0.98 : 0.02 → BIMOTOR	0.36 : 0.64 → MONOMOTOR	MONOMOTOR	OPERAÇÃO POLICIAL	TURBOEIXO	ENTRE 2250 E 5700 KG	ADMINISTRAÇÃO DIRETA	INCIDENTE
24	0.46 : 0.54 → MONOMOTOR	0.76 : 0.24 → BIMOTOR	MONOMOTOR	TÁXI AÉREO	PISTÃO	ATÉ 2250 KG	TÁXI AÉREO	INCIDENTE
25	0.00 : 1.00 → MONOMOTOR	0.00 : 1.00 → MONOMOTOR	MONOMOTOR	VOO EXPERIMENTAL	PISTÃO	ATÉ 2250 KG	EXPERIMENTAL	ACIDENTE

Caso ocorra um acidente aeronáutico em que a aeronave esteja operando em voo regular (aeronave comercial), com peso máximo de decolagem superior a 5700 KG e o motor é a jato, os dois classificadores previram com probabilidade de 100% de a aeronave ser bimotor (tabela 5, linha 1).

Ainda na tabela 5, linha 11, o classificador Naive bayes e Random forest, previram monomotor, com 100% e 93% de probabilidade, respectivamente. Porém, a variável alvo está como bimotor. Isto ocorre, pois pode haver ruído de classe.

Na linha 18, tabela 6. Caso uma aeronave efetuando um voo privado, com peso máximo de decolagem até 2250KG, sofra um incidente, sendo seu registro particular e o motor a pistão, a probabilidade de não ser monomotor é de 2% e por exclusão, ser monomotor é de 98% para o classificador Naive bayes. Já para o classificador Random foreste, a probabilidade de ser monomotor é de 67%. Visto que a variável alvo é monomotor.

Os resultados acima expostos têm como objetivo aprimorar trabalhos futuros que possam contribuir para a Segurança na Aviação Civil Brasileira. Com o advento da Mineração de Dados e Aprendizagem de Máquina, a comunidade aeronáutica tem ferramentas para gerenciar o risco e dessa forma mapear as ocorrências aeronáuticas futuras através de dados históricos registrados em um banco de dados oficial.

Aprendizagem de Máquina é uma aplicação de modelos genéricos para ajustar ou classificar algoritmos à Mineração de Dados preditivas. A ênfase na mineração de dados é a precisão da previsão, em vez de possuir um entendimento interpretável e claro da previsão (STOLZER, HALFORD & GOGLIA, 2011, p. 276).

7. Links

Vídeo

https://pucminas.instructure.com/accounts/1/external_tools/33?launch_type=global_navigation

DADOS

<https://onedrive.live.com/?id=3923EC75FA233431%21510&cid=3923EC75FA233431>

REFERÊNCIAS

Agência Nacional de Aviação Civil (ANAC). <https://www.anac.gov.br/assuntos/dados-e-estatisticas/aeronaves>. Acessado em: 21 fev. 2020.

ANJOS, C.S. **Extração e Seleção de Atributos**. Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais, Inconfidentes, 2019.

Bari, A., Chaoumaed, M., Jung, T. (2019). **Análise Preditiva para Leigos**. Alta Books.

BRASIL. Comando da Aeronáutica. Centro de Investigação e Prevenção de Acidentes Aeronáuticos. Manual de Prevenção do SIPAER - **MCA 3-3**.

BRASIL. Comando da Aeronáutica. Centro de Investigação e Prevenção de Acidentes Aeronáuticos. Centro de Investigação e Prevenção de Acidentes Aeronáuticos. **NSCA 3-3**. 2017.

BRASIL. Comando da Aeronáutica. Centro de Investigação e Prevenção de Acidentes Aeronáuticos. Manual de Investigação do SIPAER - **MCA 3-6**.

FOSTER PROVOST & TOM FACETT. **Data Science para Negócios**. Brasil: Alta Books, 2013.

Governo Federal. **Portal de Dados Abertos**: <http://www.dados.gov.br/dataset/ocorrencias-aeronauticas-da-aviacao-civil-brasileira>.

Acessado em: 20 jan. 2020.

Orange Data Mining: <https://orange.biolab.si/widget-catalog>. Acessado em: 01 fev. 2020.

NETTO, O. P. **Um Filtro Interativo Utilizando Árvores de Decisão**. Universidade de São Paulo, 2013.

LEE, D. H. **Seleção de atributos importantes para a extração de conhecimentos de bases de dados**. Universidade de São Paulo, 2005.

Instituto Brasileiro de Aviação. **Anuário Brasileiro da Aviação Civil**, 2019.

STOLZER, A.J., Halford, C.D., & Goglia, J.J. (2011). **Sistemas de Gerenciamento da Segurança Operacional na Aviação**. São Paulo, SP: DCA-BR