



Universidade do Minho

Departamento de Informática

Mestrado [Integrado] em Engenharia Informática

Mestrado em Matemática e Computação

Dados e Aprendizagem Automática

1º Ano, 1º Semestre

Ano letivo 2021/2022

Trabalho Prático de Grupo

Novembro, 2021

Tema	Conceção e otimização de modelos de <i>Machine Learning</i> .
Objetivos de Aprendizagem	Com a realização deste trabalho prático pretende-se sensibilizar e motivar os alunos para a conceção e desenvolvimento de um projeto de <i>Machine Learning</i> utilizando os modelos de aprendizagem abordados ao longo do semestre.
Enunciado	<p>A modelação do fluxo de tráfego rodoviário é um conhecido problema de características estocásticas, não-lineares. Tem, contudo, aparecido na literatura um conjunto de modelos que demonstram um potencial assinalável neste tipo de previsões. Com isso em consideração, foi construído um <i>dataset</i> que contém dados referentes ao tráfego de veículos numa cidade portuguesa durante um período superior a 1 ano. O objetivo deste trabalho passa por, entre outros, desenvolver modelos de <i>Machine Learning</i> capazes de prever o fluxo de tráfego rodoviário, numa determinada hora, na referida cidade.</p> <p>Este enunciado prático engloba duas fases.</p> <p>A <u>1ª fase da componente prática de avaliação em grupo</u> incidirá sobre as seguintes tarefas:</p> <ul style="list-style-type: none">• Consultar, analisar e selecionar um <i>dataset</i> de entre os que estão acessíveis a partir de fontes como, por exemplo, o Google Dataset Search ou Kaggle;• Para além do <i>dataset</i> selecionado no ponto anterior, os grupos deverão também trabalhar o <i>dataset</i> disponível em https://www.kaggle.com/c/tpsbsd2019:<ul style="list-style-type: none">○ O link anterior redireciona para a plataforma <i>Kaggle</i> onde foi criada uma competição. O <i>dataset</i> a utilizar na competição, assim como todos os detalhes do funcionamento da mesma, estão disponíveis no referido link;○ O primeiro passo consiste em aceder à plataforma <i>Kaggle</i>, utilizando o seguinte link para se inscreverem na competição: https://www.kaggle.com/t/0294f9d234774d24ba3e16803a178c05○ Devem, de seguida, formar equipas com os restantes elementos do grupo de trabalho. O nome da equipa deverá seguir o formato GRUPO_<CURSO>_<X> onde <CURSO> corresponde ao curso de mestrado (MMC, MEI ou MIEI) e <X> ao número do grupo. Não poderão efetuar submissões na plataforma <i>Kaggle</i> enquanto o grupo se apresentar incompleto.• Explorar, analisar e preparar ambos os <i>datasets</i>, procurando extrair conhecimento relevante no contexto dos problemas em questão.

A 2ª fase da componente prática de avaliação em grupo incidirá sobre as seguintes tarefas:

- Conceção e otimização de modelos de *Machine Learning* para ambos os problemas, i.e., para o *dataset* escolhido pelo grupo de trabalho e para o *dataset* da competição:
 - Em relação a este último, deverão submeter os resultados obtidos na plataforma *Kaggle* de forma a obter a *accuracy* do modelo;
 - Existe um **limite diário de 3 submissões válidas** pelo que deverão procurar começar as submissões assim que possível. A competição encerra no dia 17 de dezembro de 2021.
- Obtenção e análise crítica de resultados;
- Interpretação dos resultados adquiridos e definição da sua utilidade no contexto dos problemas subjacentes aos *datasets* trabalhados. Determinar e explicitar os resultados mais relevantes.

Os resultados obtidos deverão ser objeto de um relatório, limitado a 20 páginas, que apresente, entre outros:

- Quais os domínios a tratar, quais os objetivos e como se propõe a atingi-los;
- Qual a metodologia seguida e como foi aplicada;
- Descrição e exploração detalhada de ambos os *datasets* e de todo e qualquer tratamento efetuado;
- Descrição dos modelos desenvolvidos, quais as suas características, como e sobre que parâmetros foi realizado o *tuning* do modelo, características do treino, entre outros detalhes que seja oportuno fornecer;
- Sumário dos resultados obtidos e respetiva análise crítica;
- Apresentação de sugestões e recomendações após análise dos resultados obtidos e dos modelos desenvolvidos.

Todo o processo deverá ser acompanhado de exemplos e indicações que permitam reproduzir todos os passos realizados assim como os resultados obtidos.

Entrega e Avaliação

Durante o período de aulas do dia 2 de dezembro de 2021 decorrerá a avaliação da 1ª fase da componente prática de avaliação em grupo. No referido dia será feito um checkpoint ao trabalho desenvolvido pelos grupos de trabalho, devendo cada grupo utilizar os meios que considerar mais adequados para demonstrar os resultados obtidos. Este checkpoint decorrerá de forma remota, sendo utilizada a ferramenta *Blackboard Collaborate Ultra* da plataforma de e-learning da Universidade do Minho.

Na semana de 3-7 de janeiro de 2022 decorrerão as sessões de apresentação do trabalho desenvolvido em ambas as fases. Os grupos de trabalho deverão escolher o *slot* desejado para realização da apresentação, sendo que esses *slots* serão disponibilizados nas próximas semanas. Cada grupo disporá de 10 minutos para realizar a apresentação, utilizando os meios que considerar mais adequados. A apresentação decorrerá de forma remota, sendo utilizada a ferramenta *Blackboard Collaborate Ultra* da plataforma de e-learning da Universidade do Minho.

O relatório, assim como os restantes elementos produzidos, deverão ser compactados num único ficheiro zip que deverá ser submetido, por um elemento do grupo, na plataforma de e-learning da Universidade do Minho (em “*Conteúdo/Instrumentos de Avaliação em Grupo/Submissão TPG*”). As submissões deverão ser realizadas até ao final do dia 3 de janeiro de 2022.

Código de Conduta

Os participantes do presente trabalho académico declaram ter atuado com integridade e confirmam que não recorreram à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração. Mais declaram que conhecem e respeitaram o Código de Conduta Ética da Universidade do Minho.

**Referências
Bibliográficas**

Além do material disponibilizado nas aulas, aconselha-se a consulta de fontes como:

Machine Learning. T. Michell, McGraw Hill, ISBN ISBN: 978-1259096952, 2017.

Introduction to Machine Learning. Alpaydin, E. ISBN: 978-0-262-02818-9. Published by The MIT Press, 2014.

Computational Intelligence: An Introduction. Engelbrecht A., Wiley & Sons. 2nd Edition, ISBN: 978-0470035610, 2007.

The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Hastie, T., R. Tibshirani, J. Friedman, 12nd Edition, Springer, ISBN: 978-0387848570, 2016.

Machine Learning: A Probabilistic Perspective. K.P. Murphy, 4th Edition, The MIT Press, ISBN: 978-0262018029, 2012.