%\title{LaTeX Portrait Poster Template}

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

% a0poster Portrait Poster

% LaTeX Template

% Version 1.0 (22/06/13)

%

% The a0poster class was created by:

% Gerlinde Kettl and Matthias Weiser (tex@kettl.de)

%

% This template has been downloaded from:

% http://www.LaTeXTemplates.com

%

% License:

% CC BY-NC-SA 3.0 (http://creativecommons.org/licenses/by-nc-sa/3.0/)

%

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

%----------------------------------------------------------------------------------------

% PACKAGES AND OTHER DOCUMENT CONFIGURATIONS

%----------------------------------------------------------------------------------------

\documentclass[a0,portrait]{a0poster}

\usepackage{multicol} % This is so we can have multiple columns of text side-by-side

\columnsep=100pt % This is the amount of white space between the columns in the poster

\columnseprule=3pt % This is the thickness of the black line between the columns in the poster

\usepackage[svgnames]{xcolor} % Specify colors by their 'svgnames', for a full list of all colors available see here: http://www.latextemplates.com/svgnames-colors

\usepackage{times} % Use the times font

%\usepackage{palatino} % Uncomment to use the Palatino font

\usepackage{graphicx} % Required for including images

\graphicspath{{figures/}} % Location of the graphics files

\usepackage{booktabs} % Top and bottom rules for table

\usepackage[font=small,labelfont=bf]{caption} % Required for specifying captions to tables and figures

\usepackage{amsfonts, amsmath, amsthm, amssymb} % For math fonts, symbols and environments

\usepackage{wrapfig} % Allows wrapping text around tables and figures

\begin{document}

%----------------------------------------------------------------------------------------

% POSTER HEADER

%----------------------------------------------------------------------------------------

% The header is divided into two boxes:

% The first is 75% wide and houses the title, subtitle, names, university/organization and contact information

% The second is 25% wide and houses a logo for your university/organization or a photo of you

% The widths of these boxes can be easily edited to accommodate your content as you see fit

\begin{minipage}[b]{0.75\linewidth}

\VeryHuge \color{NavyBlue} \textbf{Drug Use Level Prediction} \color{Black}\\ % Title

\huge\textit{Predict drug use level from individual's background and personality traits}\\[1.4cm] % Subtitle

\huge \textbf{Hu Tang}\\[0.5cm] % Author(s)

\huge Imperial College London\\[0.4cm] %University/organization

\Large \texttt The contents of this work and the associated code are my own unless otherwise stated.\\

\end{minipage}

%

\begin{minipage}[b]{0.25\linewidth}

\includegraphics[width=20cm]{Imperial-College-London.jpg}\

\end{minipage}

\vspace{1cm} % A bit of extra whitespace between the header and poster content

%----------------------------------------------------------------------------------------

\begin{multicols}{3} % This is how many columns your poster will be broken into, a portrait poster is generally split into 2 columns

%----------------------------------------------------------------------------------------

% ABSTRACT

%----------------------------------------------------------------------------------------

\color{Navy} % Navy color for the abstract

\begin{abstract}

Anyone, at any age or any stage of their life can have a substance use problem, including substances like alcohol, cigarettes and illegal drugs. World Health Organization claims that about 31 million persons have drug use disorders, and almost 11 million people inject drugs, of which 1.3 million are living with HIV, 5.5 million with hepatitis C, and 1 million with both HIV and hepatitis C.[WTO] It is useful if we could predict individual's drug use level, separate them into two groups and treat differently to prevent them from future drug uses.

\end{abstract}

%----------------------------------------------------------------------------------------

% INTRODUCTION

%----------------------------------------------------------------------------------------

\color{Black} % SaddleBrown color for the introduction

\section\*{Introduction}

We will use a dataset that contains:\\

\textbf{Predictors}:

\begin{itemize}

\item individual’s background : agegroup, gender, education, country, ethnicity

\item scores for personality traits : neuroticism, extraversion, opentoexperience, agreeableness, conscientiousness, impulsiveness, sensation

\item legal substances: caffeine, chocolate, nicotine, alcohol)

\end{itemize}

\textbf{Outcomes}:

\begin{itemize}

\item Severity : score of the severity of drug consumption.

\item Uselevel : "high" or "low" use level of drug consumption.

\end{itemize}

We will predict Uselevel of the individuals based on the predictors.

%----------------------------------------------------------------------------------------

% GEOLOGY

%----------------------------------------------------------------------------------------

\color{Black} % DarkSlateGray color for the rest of the content

\section\*{Exploratory data analysis}

\begin{center}\vspace{1cm}

\includegraphics[width=0.8\linewidth]{2.png}

\captionof{figure}{\color{Blue} It is clear from the plot that there is a boundary that separate low and high use levels: opentoexperience and use level has a strong positive correlation, while conscientiousness and use level has a strong negative correlation.}

\end{center}%\vspace{1cm}

\begin{center}\vspace{1cm}

\includegraphics[width=0.8\linewidth]{3.png}

\captionof{figure}{\color{Blue} The plot illustrates that use level and opentoexpeirence has a strong positive correlation.}

\end{center}%\vspace{1cm}

\begin{center}\vspace{1cm}

\includegraphics[width=0.8\linewidth]{4.png}

\captionof{figure}{\color{Blue} The plot shows that younger individuals are more likely to have a high use level, and female tends to have a low use level comparing to male.}

\end{center}%\vspace{1cm}

\begin{center}\vspace{1cm}

\includegraphics[width=0.8\linewidth]{5.png}

\captionof{figure}{\color{Blue} It is obvious from the graph that male generally have higher severity than female, and individuals in UK has lower severity than individuals in other countries.}

\end{center}%\vspace{1cm}

\begin{center}\vspace{1cm}

\includegraphics[width=0.8\linewidth]{6.png}

\captionof{figure}{\color{Blue} The plot tells us that younger people are more likely have higher severity. There is a strong negative correlation between severity and age group.}

\end{center}%\vspace{1cm}

From the plots, we found that there are some predictors that are strongly correlated with use level.

%----------------------------------------------------------------------------------------

% GEOTHERMAL DATA

%----------------------------------------------------------------------------------------

\section\*{Machine Learning}

\subsection\*{Logistic regression model}

I used logistic regression to build a classifier that predicts if an individual’s substance use level will be ‘high’ or ‘low’ based on the predictors.

\begin{center}\vspace{1cm}

\includegraphics[width=1.0\linewidth]{roc.png}

\captionof{figure}{\color{Blue} The ROC curve is high above $y=x$, which means it is a good classifier. Also a calculated AUC value of 0.93 also suggests that the classifier performs well.}

\end{center}\vspace{1cm}

I then used 10-fold Cross Validation to estimate the accuracy I would see if I use the model on a new data set.

\begin{center}\vspace{1cm}

\includegraphics[width=1.0\linewidth]{accuracy.png}

\captionof{figure}{\color{Blue} If I use the model on a new data set, I will get an accuracy of 0.86, which is high}

\end{center}\vspace{1cm}

%------------------------------------------------

\subsection\*{Ensemble learning}

I will try another model to see if I can achieve a higher accuracy than logistic regression model: use three models on training data: KNN, Random Forest and SVM. For each of these models, I tune the hyper parameter respectively.\\

Finally, I use these models to predict the test data, and ensemble the three predictions with majority vote.\\

Confusion matrix is generated:

\begin{center}\vspace{1cm}

\makegapedcells

\begin{tabular}{cc|cc}

\multicolumn{2}{c}{}

& \multicolumn{2}{c}{Actual} \\

& & High & Low \\

\cline{2-4}

\multirow{\rotatebox[origin=c]{90}{Predict}}

& High & 190 & 24 \\

& Low & 27 & 136 \\

\cline{2-4}

\end{tabular}

\end{center}\vspace{1cm}

This time the accuracy is 0.865, higher than logistic regression model.

\section\*{Predictor Importance}

It is essential to say which predictors are important in predicting illegal drug use. For this study, I will discuss the predictor importance in two aspects:

\subsection\*{Logistic regression summary}

We perform significance test on predictors with significance level of 5\%. By comparing the p-values of predictors with 5\%, we have enough evidence to say that age group, gender, education, ethnicity, extraversion, opentoexperience, conscientiousness, sensation, chocolate, nicotine and alcohol has association with the use level. We consider them as important predictors.

\subsection\*{Random forest variable importance}

\begin{center}\vspace{1cm}

\includegraphics[width=1.0\linewidth]{importance.png}

\captionof{figure}{\color{Blue} Mean decrease Gini are important index for variable importance. Variables with a large mean decrease in accuracy and mean decrease in Gini are more important for classification of the data. From the graph we can see that, nicotine, country, sensation, opentoexperience, conscientiousness, age group, gender, education and neuroticism are important predictors.}

\end{center}\vspace{1cm}

In conclusion, nicotine use, sensation, opentoexperience and conscientiousness personality traits , age group and gender are considered as important predictors in both aspects.

\section\*{Conclusions}

We found that it is accurate to predict individual's drug use level just from individual's background and personality traits, and it will be efficient to separate them into two groups and treat differently. Also, nicotine use, sensation, opentoexperience and conscientiousness personality traits , age group and gender are important predictors in predicting drug use level.

\color{Black} % Set the color back to DarkSlateGray for the rest of the content

%----------------------------------------------------------------------------------------

% FORTHCOMING RESEARCH

%----------------------------------------------------------------------------------------

%----------------------------------------------------------------------------------------

% REFERENCES

%----------------------------------------------------------------------------------------

\bibliography{Reference}

\bibitem{facts}

“Facts and Figures.” \textit{World Health Organization}, 17 Aug. 2018, www.who.int/substance\\_abuse/facts/en/.

\bibliographystyle{chicago}

%----------------------------------------------------------------------------------------

\end{multicols}

\end{document}