MSc. in Computing
Practicum Approval Form

Section 1: Student Details

Project Title:	Explainable AI for the Analysis of the Structure-Odour-Relationship
Student ID:	10377593
Student name:	Denis Kealy
Student email	denis.kealy2@mail.dcu.ie
Chosen major:	Data Analytics
Supervisor	Martin Crane / Tim Downing
Date of Submission	10/11/17

Section 2: About your Practicum

What is the topic of your proposed practicum? (100 words)

The aim of my project is to develop an explainable machine learning system or an interface for an existing machine learning framework which gives a description of the machine's decision process along with the output. Machine Learning programs have been proven to have a high odour-perception prediction rate when shown novel odorant molecules (molecules not in the training set). If we can approach this level of prediction in our system - the outputted description of the machine decision process (given as a set of physical descriptors) would be very useful for understanding how machines are classifying these molecules to achieve such high prediction rates. This approach black boxes the neural processing components of the problem.

Please provide details of the papers you have read on this topic (details of 5 papers expected).

1. **Predicting human olfactory perception from chemical features of odour molecules**
   - It is still not possible to predict whether a given molecule will have a perceived odour or what olfactory percept it will produce. We therefore organized the crowd-sourced DREAM Olfaction Prediction Challenge. Using a large olfactory psychophysical data set, teams developed machine-learning algorithms to predict sensory attributes of molecules based on their chemoinformatic features. The resulting models accurately predicted odour intensity and pleasantness and also

successfully predicted 8 among 19 rated semantic descriptors (“garlic,” “fish,” “sweet,” “fruit,” “burnt,” “spices,” “flower,” and “sour”). Regularized linear models performed nearly as well as random forest–based ones, with a predictive accuracy that closely approaches a key theoretical limit. These models help to predict the perceptual qualities of virtually any molecule with high accuracy and also reverse-engineer the smell of a molecule. – ***Science Mag - 24 Feb 2017: Vol. 355, Issue 6327, pp. 820-826 DOI: 10.1126/science.aal2014***
- http://science.sciencemag.org.dcu.idm.oclc.org/content/355/6327/820.full

2. **Understanding the Odour Spaces: A Step towards Solving Olfactory Stimulus-Percept Problem -** Odours are highly complex, relying on hundreds of receptors, and people are known to disagree in their linguistic descriptions of smells. It is partly due to these facts that, it is very hard to map the domain of odour molecules or their structure to that of perceptual representations, a problem that has been referred to as the Structure-Odour-Relationship. We collected a number of diverse open domain databases of odour molecules having unorganised perceptual descriptors, and developed a graphical method to find the similarity between perceptual descriptors; which is intuitive and can be used to identify perceptual classes. We then separately projected the physico-chemical and perceptual features of these molecules in a non-linear dimension and clustered the similar molecules. We found a significant overlap between the spatial positioning of the clustered molecules in the physico-chemical and perceptual spaces. We also developed a statistical method of predicting the perceptual qualities of a novel molecule using its physico-chemical properties with high receiver operating characteristics(ROC). **- Published: October 20, 2015 - http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0141263**
3. **DARPA-BAA-16-53 - Research Funding Announcement - Explainable Artificial Intelligence (XAI) -** This document details some of the problem areas and challenges concerned with solving this problem. This is where I initially came across the concept of Explainable AI - *August 10, 2016* **https://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf**
4. **Making machine learning models interpretable -** Data of different levels of complexity and of ever growing diversity of characteristics are the raw materials that machine learning practitioners try to model using their wide palette of methods and tools. The obtained models are meant to be a synthetic representation of the available, observed data that captures some of their intrinsic regularities or patterns. Therefore, the use of machine learning techniques for data analysis can be understood as a problem of pattern recognition or, more informally, of knowledge discovery and data mining. There exists a gap, though, between data modeling and knowledge extraction. Models, depending on the machine learning techniques employed, can be described in diverse ways but, in order to consider that some knowledge has been achieved from their description, we must take into account the human cognitive factor that any knowledge extraction process entails. These models as such can be rendered powerless unless they can be interpreted, and the process of human interpretation follows rules that go well beyond technical prowess. For this reason, interpretability is a paramount quality that machine learning methods should aim to achieve if they are to be applied in practice. - *ESANN 2012 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. 25-27 April 2012 - ISBN 978-2-87419-049-0. - https://www.elen.ucl.ac.be/Proceedings/esann/esannpdf/es2012-7.pdf*
5. **"Why Should I Trust You?": Explaining the Predictions of Any Classifier -** Despite

widespread adoption, machine learning models remain mostly black boxes. Understanding the reasons behind predictions is, however, quite important in assessing trust, which is fundamental if one plans to take action based on a prediction, or when choosing whether to deploy a new model. Such understanding also provides insights into the model, which can be used to transform an untrustworthy model or prediction into a trustworthy one. In this work, we propose LIME, a novel explanation technique that explains the predictions of any classifier in an interpretable and faithful manner, by learning an interpretable model locally around the prediction. We also propose a method to explain models by presenting representative individual predictions and their explanations in a non-redundant way, framing the task as a submodular optimization problem. We demonstrate the flexibility of these methods by explaining different models for text (e.g. random forests) and image classification (e.g. neural networks). We show the utility of explanations via novel experiments, both simulated and with human subjects, on various scenarios that require trust: deciding if one should trust a prediction, choosing between models, improving an untrustworthy classifier, and identifying why a classifier should not be trusted. **Paper -** https://arxiv.org/abs/1602.04938. **Github** - https://github.com/marcotcr/lime

How does your proposal relate to existing work on this topic described in these papers? (200 words)

Paper 1 closely resembles the process of my own investigation. I intend to use their methods as an outline for my investigation with the addition of explainable machine model outputs. I would ideally like to make use of the data set used in this experiment but should it not be possible to obtain the raw data set used in this paper we can take the approach of Paper 2.

Paper 2 uses a set of 5 publicly available online databases with contain data on molecules as they relate to their human perceived scent. In this paper, they clustered the molecules by certain physio-chemical and perceptual features and found that a classification of perceptual smells (odour-space) could be drawn these results. Finally, they also, as in Paper 1, trained a machine learning program to predict the perceptual qualities of novel molecules.

Paper 3 is a new paradigm that has been suggested for learning machines by which they explain the rules that they have developed to the human user. Techniques such as this could be used to help formulate or pinpoint viable theories of Olfaction.

Paper 4 is a 2012 paper that discusses how to interpret machine models by various means, most notably by reducing dimensionality. This relates to my work as reducing the dimensionality is essential to making models interpretable, as my model will have to be.

Paper 5 is a computer science paper that attempts to address the problem of interpretability in machine learning models. They have released the source code for their implementation of an explainable AI framework for any classifier. This code will serve as a starting point for my own development or it may be entirely sufficient for my analysis purposes.

How will you explore these questions? (Please address the following points. Note that three or four sentences on each will suffice.)

- What software and programming environment will you use?

LIME – https://github.com/marcotcr/lime

Python with PyCharm (JetBrains IDE)

- What coding/development will you do?

I will attempt to configure LIME to work with an existing machine learning algorithm such as random-forest classifiers, which were the most successful machine learning algorithms used in 2017 paper (Paper 1) placing first in the ranking of predictive accuracy.

- What data will be used for your investigations?

Online databases such as:

- GoodScents
- SuperScent
- Flavornet
- and Sigma-Aldrich

Unpublished data set of 480 molecules and their perceived odour, by 50 participants – Paper [Link]

Physical Descriptor sets

- Is this data currently available, it not, where will it come from?

Data publically available online – with the possibility of obtaining experimental data from researchers at The Rockefeller University, NY, USA.

- What experiments do you expect to run?

After developing an explainable machine learning program I intend to test its predictive powers against a set of unseen molecules. I will perform statistical analysis of the Structure-Odour-Relationship to hopefully produce descriptive outputs that can help identify viable theories of Olfaction.

- What output do you expect to gather?

This system is expected to produce an output of a predicted perception of scent, given a physical description of a molecule (molecule is input as a set of physical descriptors). As well as a prediction our system will attempt produce an explanation or justification for it's output.

- How will the results be evaluated?

I will use a similar method of evaluation as Paper 1 & 2. Data visualisations as well the source code for my program will be released as a part of my work so that experts in the fields of Chemistry, Bioinformatics and Machine Learning can create and perform their own experiments. Experts can use their domain knowledge of the problem to choose and manipulate the chemical descriptors as necessary.