BINUS University

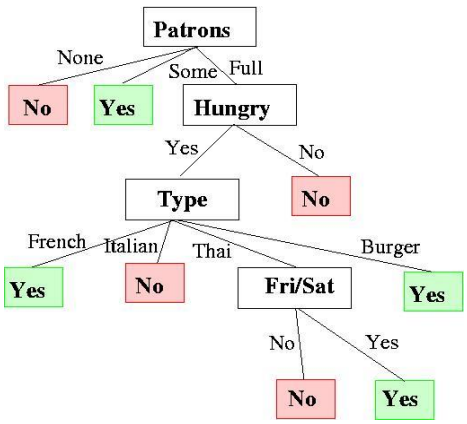
Academic Career: ***Undergraduate / ~~Master~~ / ~~Doctoral~~ *)***			**Class Program:** ***~~International~~/Regular/~~Smart Program~~/~~Global Class~~*)***		
☐ **Mid Exam** ☑ **Final Exam** ☐ **Short Term Exam** ☐ **Others Exam : _____**			**Term : ~~Odd~~/Even/~~Short~~ *)**		
☑ **Kemanggisan** ☑ **Alam Sutera** ☐ **Bekasi** ☐ **Senayan** ☐ **Bandung** ☐ **Malang**			**Academic Year:** **2021 / 2022**		
Faculty / Dept.	:	School of Computer Science	Deadline	Day / Date :	Kamis / 10 Februari 2022
				Time :	13:00
Code - Course	:	COMP6639001-ARTIFICIAL INTELLIGENCE	Class	:	All Classes
Lecturer	:	Team	Exam Type	:	Online
*) *Strikethrough the unnecessary items*					
The penalty for CHEATING is DROP OUT!!!					

Learning Outcomes:

LO 5 : Apply various techniques to an agent when acting under certainty
LO 6 : Apply various AI algorithms to solve the problems

I. ESSAY (100 %)

1. **[LO 5 & LO 6, 20 points]** Observe thoroughly the restaurant_dat.csv dataset (you can find a similar dataset in the textbook AIMA 3rd p. 700 Fig. 18.3) and the corresponding decision tree shown in the below figure. Using **the entropy E and information gain IG**, show by hand calculation how the attributes such as Patrons (Pat), Hungry (Hun), Type (Type), and Fri/Sat (FriSat) are chosen as a root node or decision nodes respectively. Based on your comprehensive E and IG calculations, could we also get other decision tree representations ?. Please give your analysis on this matter.



2. **[LO 5 & LO 6, 20 points]** A Data Scientist, you are supposed to build a multivariate linear regression (MLR) model that might be used to predict the Stock Price Index (dependent variable) based on two independent variables namely Interest Rate and Unemployment Rate as they are shown in the **mlr_dat.csv** dataset. **Find**

Verified by,

Hidayaturrahman (D6423) and sent to Program on JAN 17, 2022

the intercept *b* and the coefficient W_i analytically by consulting the aforementioned problem to the textbook AIMA 3rd Edition pp. 718-721. Compare and show side by side to see how close your MLR model prediction with the actual one (ground truth).

3. **[LO 5 & LO 6, 20 points]** Construct a support vector machine SVM (**by hand calculating it's hyperparameters W_i and *b***) that computes the AND function. Use values +1 and -1 (instead of 1 and 0) for the target output, so that the four input points and the corresponding target outputs look like ([0, 0], -1); ([1, 0], -1); ([0, 1], -1); and ([1, 1], 1]. Draw those four input points along with the **decision boundary function (*W*X +b*)** in the Euclidean space. What is the margin?. Check also the correctness of your SVM model with the test input points [1.5, 0.2] and [0.8, 02]. Please consult to AIMA 3rd Edition pp.744-748.

4. **[LO 5 & LO 6, 20 points]** Train by hand a single neuron using perceptron learning rules on the training set given below. Assume that all initial weights including the bias of the neuron are zeros (0). Show the set of weights including the bias at the end of each iteration. Use learning rate =1 and the **threshold function $h_w(X)$** : threshold(net) = 1 if net >= 0; threshold(net) = 0 otherwise (pls check it out, textbook AIMA 3rd Edition pp. 723-724). Apply the examples in the given order and stop the iteration by the time when you find the patterns are all correctly classified. Is the training set linearly separable, give your comment on it.

Expl. No	Input	Output
1	1 0 0	1
2	0 1 1	0
3	1 1 0	1
4	1 1 1	0
5	0 0 1	0
6	1 0 1	1

5. **[LO 5 & LO 6, 20 points]** Dropping stopwords and stemming/lemmatization process on a corpus (a set of text based documents) produce the following document space:

"alone, animation, best, cold, enjoy, ever, filled, film, he, horror, i, love, most, movie, night, one, outside, overrated, put, plot, really, seen, story, scary, suspense, toy, twist, unexpected, watching".

If the corresponding set of text based documents consist of:

D1 = I enjoy watching movies when it's cold outside;
D2 = Toy story is the best animation movie ever;
D3 = Watching horror movies alone at night is really scary;
D4 = He loves film filled with suspense and unexpected plot twists
D5 = This is one of the most overrated movie I've ever seen.

and if the query document Q contains key terms ***watching best animation movies***, perform hand calculation to determine the rank of every single document above if the retrieval process is performed using BM25 approach. Please use k_1 = 1.5, b = 0,75, and to avoid IDF with negative values please slightly modify IDF equation (AIMA 3rd Edition, p.868-869) to become: $IDF(q_i)=\log_e(1+(N-DF(q_i)+0.5)/(DF(q_i)+0.5))$ where ***N*** is the number of documents and $DF(q_i)$ is the number of documents in the corpus that contain term q_i.

Verified by,

Hidayaturrahman (D6423) and sent to Program on JAN 17, 2022