3BA2 Tutorial 2

Decision Trees

1. In this question we consider the use of Machine Learning Induction techniques in the design of an E-mail filtering system. The objective of the system is to rank incoming messages as Important or Normal depending on the attributes of some messages.
An analysis of the problem has shown that three key predictive features of the importance of a message are:

   Domain: The origin of the message; can be Internal, Ireland or the rest if the World.

   Size: Can be Small, Medium or Large.

   Type: Whether the message was addressed to a Mailing List or an Individual.

   In the Table below there are 10 examples of email messages described according to these features. Show how a minimal decision tree to classify email messages can be built by Induction from these examples.

Message	Domain	Size	Type	Rank
No.1	Internal	Small	Personal	**Important**
No.2	Internal	Medium	Mailing List	**Normal**
No.3	Internal	Small	Personal	**Important**
No.4	World	Small	Personal	**Important**
No.5	World	Large	Mailing List	**Normal**
No.6	IE	Small	Mailing List	**Normal**
No.7	IE	Small	Personal	**Important**
No.8	World	Large	Mailing List	**Normal**
No.9	IE	Large	Personal	**Normal**
No.10	IE	Medium	Personal	**Normal**

2. Build a decision tree from the following data that will classify an individual as McDonnell or Giles.

	Height	Hair	Eyes	Clan
Sean	short	blond	blue	**McD**
Mike	tall	blond	brown	**Giles**
Paddy	tall	red	blue	**McD**
Mike Óg	short	dark	blue	**Giles**
Colm	tall	dark	blue	**Giles**
Liam	tall	blond	blue	**McD**
Johnny	tall	dark	brown	**Giles**
Cóilín	short	blond	brown	**Giles**