

## ***Guida per un corretto utilizzo della funzionalità di upload di dataset nel client***

Per effettuare predizioni viene fornita all'utente la possibilità di caricare sul server dataset presenti sulla macchina in uso, indipendentemente se il client stia venendo eseguito da un browser o come applicazione desktop.

I dataset che l'utente può inviare devono essere in specifici formati e presentare una struttura appropriata per il tipo di file per permettere al server di generare correttamente l'albero di decisione ed effettuare predizioni esatte.

*N.B.: Non si fornisce alcuna garanzia sul corretto funzionamento del meccanismo di parsing del dataset e di generazione dell'albero di decisione nel caso in cui vengano inviati al server file non conformi alle direttive che seguono.*

### **Formati supportati**

- .txt** (file di testo – vedasi: [http://it.wikipedia.org/wiki/File\\_di\\_testo](http://it.wikipedia.org/wiki/File_di_testo))
- .xls** (fogli di Microsoft Excel – vedasi: [http://it.wikipedia.org/wiki/Microsoft\\_Excel](http://it.wikipedia.org/wiki/Microsoft_Excel))
- .arff** (dataset – vedasi: <http://it.wikipedia.org/wiki/ARFF> e <http://www.cs.waikato.ac.nz/ml/weka/arff.html> )

Per ogni formato è prevista una precisa strutturazione del file.  
Alcuni esempi sono presenti nella relativa cartella per maggiore chiarezza.

### **Formato .txt**

La prima riga del file deve contenere i **nomi degli attributi** separati da *spazio*.

La seconda riga del file deve contenere i rispettivi **tipi degli attributi** separati da *spazio*. I tipi consentiti sono *String* e *Float*.

Dalla terza in riga in poi si possono inserire i valori delle **tuple**: scrivere una tupla per riga, separando i valori con la *virgola* (senza *spazio*).

### **Formato .xls**

L'elaborazione del dataset sarà effettuata solo sul primo foglio di calcolo del file. Eventuali altri fogli saranno ignorati.

La prima riga del file deve contenere i **nomi degli attributi**, ciascuno in una cella.

La seconda riga del file deve contenere i rispettivi **tipi degli attributi**, ciascuno in una cella. I tipi consentiti sono *String* e *Float*.

Dalla terza in riga in poi si possono inserire i valori delle **tuple**: scrivere una tupla per riga, inserendo ogni valore nella rispettiva cella.

### **Formato .arff**

I più comuni dataset in formato .arff saranno correttamente elaborati dal server, tuttavia si rende noto che non sono supportate tutte le caratteristiche che la notazione ufficiale dei file .arff offre (ad esempio, non è supportato il carattere jolly ? all'interno delle tuple).

Di seguito vi è una indicazione di come può essere strutturato il file .arff affinché venga correttamente elaborato dal server.

Le prime righe del file .arff possono contenere commenti e informazioni sul dataset. I commenti vengono segnalati scrivendo all'inizio della riga %. Essi verranno automaticamente ignorati.

I tag, da ora in avanti, sono da considerarsi non case-sensitive.

Oltre ai commenti, saranno ignorate tutte le righe del file che precedono la parte dichiarativa degli attributi (quindi, ad esempio, il tag *@relation* sarà ignorato).

Gli attributi si dichiarano in una riga seguendo questo schema:

*@attribute* **nomeattributo** *dominio*

Dove il *dominio* può essere *numeric* o *real* (non case-sensitive) oppure un dominio discreto definito dall'utente in questo modo:

*{value1, value2, value3, ..., value4}*

Successivamente è possibile definire le tuple. Per farlo bisogna scrivere in una riga il tag *@data* (e *nient'altro*) e, nelle righe successive, le tuple. I valori nelle tuple sono separati da *virgola*.