Commenti sul report:

* I report non devono essere interpretabili, ma bisogna dare chiarezza in ogni punto.
* Uniformità di grafici e caption in tutto il report (font, descrizione auto-esplicativa delle immagini)
* Figure e tabelle con testo sopra e sotto, non a lato.

Se i valori sono pochi, si commentano in una frase di testo.

* Uniformità di numerazione di sezioni e figure e tabelle
* Figure e tabelle devono leggersi (anche se si zooma)
* Coerenza delle cifre decimali in tutto il report
* Apici -> “…” aperti e chiusi (un gruppo ha fatto apici\_chiusi … apici\_chiusi)
* Tutte le tabelle e le figure vanno richiamate nel testo, possibilmente nell’ordine in cui sono mostrate.
* Esposizione finale: meglio con slide per non andare su e giù nel report
* Inserire un abstract (?)
* Inserire scientific questions (?) prima di spiegare le variabili
* Se una particolare metodologia porta a risultati schifosi, si può evitare di riportare i grafici e basta mettere che si è provato a fare qualcosa, ma che fa schifo o non comporta miglioramenti (come KNN già fatto) -> no obbligo di mettere tutte le figure per tutto quello che si spiega, ma si può anche spiegare a parole.

Lunghezza del report:

* 20 pagine sono troppe
* Introduzione e references finali
* Matrice di correlazione: bastano 2/3 cifre decimali

Introduzione:

* Da dove viene il dataset (Kaggle)
* Analisi descrittiva variabili (tabella con variabile, tipo, descrizione, unità di misura, istogrammi)

Tabelle:

* No troppi decimali, ma vanno rispettati in ogni cella della tabella
* Allineamento dei numeri a destra (il “-“ fa sfalsare l’allineamento)
* Descrizione di figure e tabelle: capire cosa c’è in tabella / figura senza leggere il paper

Non troppo sintetica.

* Unità di misura nelle tabelle
* NO TABELLE COME IMMAGINI

Figure:

* Unità di misura e descrizioni sugli assi (es: conteggio = frequency)
* Anche zoomando non si deve perdere la risoluzione dell’immagine
* Tenere i font omogenei tra le varie figure
* Linee e colori: descrizione nella caption anche se viene troppo lunga.
* Single nella legenda dei colori: vanno riportate nella caption -> legenda nella caption
* Più immagini con caption simili o uguali -> “Si veda la caption dell’immagine x”

Non ripetere la stessa caption

* Elementi grafici di un’immagine descritti nella caption (IMPORTANTE)

Descrizione iniziale delle variabili (TODO)

* Uso di istogrammi per le variabili quantitative
* Istogrammi centrati negli assi (solo se non ci sono osservazioni lontane dal centro dell’istogramma)

Alberi:

* Grafici di alberi unpruned: o senza label o non si mette perché è troppo confuso

Meglio descrivere l’albero a parole

* Classificazione su variabile categorica: spiegare i valori messi dall’algoritmo per potare l’albero (es: se la variabile assume 0 e 1, l’albero poterà su un valore di 0.5)

Risultati:

* Tabelle per confronto tra le varie tecniche: se la struttura è identica per diverse tecniche meglio accorparle in un’unica tabella e scrivere il nome della variabile sopra/sotto la tabella
* Tabella con risultati unica, senza duplicare le informazioni già scritte nel report o in altre tabelle.
* Scrittura dei valori risultanti nel testo va bene se è uno.

Inserimento MSE e basta non ha senso se non si ha il range della variabile.

Meglio un R^2 nel testo.

* Per confrontare MSE è bene usare anche delle % che dicono di quanto aumenta o diminuisce.

Bibliografia

* Aggiungere bibliografia se si fa riferimento a dei documenti o da dove si prendono varie formule (TODO)