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 (?): non necessario perché è un report.
* 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.
* Approccio top-down: prima la descrizione dei dati e poi si procede verso l’analisi.
* Colonne dataset: no parametri -> parametri sono qualcosa che si stima in statistica.

Meglio variabile, regressore, …

* Allineate al centro solo immagini, tabelle e equazioni.

Il resto va tenuto normale, senza corsivo.

* Introducendo una variabile in un certo modo deve rimanere lo stesso in tutto il report (grassetto, corsivo,…)
* Se si introduce un simbolo va utilizzato (esempio: X insieme di regressori e Y variabile risposta).
* Non saltare pagine (no spazi bianchi) e cercare di ridimensionare i margini del template (destra, sinistra, alto e basso) per guadagnare spazio.
* Confronti tra immagini nella stessa pagina per fare un confronto a colpo d’occhio
* Figure e tabelle da mostrare solo se cambia qualcosa all’interno del nuovo metodo.

Solo cose significative. (Backward)

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)
* Descrizione problema, dataset e approcci utilizzati (parlando senza entrare nella parte tecnica)

Tabelle:

* Tabelle di risultati dei vari modelli
* 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
* Unificare tabelle diverse se hanno la stessa intestazione e riportare un nome di modello diverso per far capire da dove arrivano quei valori nella tabella.

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)

Es: final model non si capisce, perché bisogna dire cosa è il modello finale

Descrizione iniziale delle variabili (TODO)

* Descrizione delle variabili più dettagliata specificando continue / discrete, input / output (elenco non tabella).
* Uso di istogrammi per le variabili quantitative -> empirical distributions / histograms
* 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)

Screenshot:

* Devono diventare tabelle

Uniformità decimali importante

* Mettere immagini e tabelle con una certa rilevanza

Meglio appendice in fondo in cui riportare diverse tabelle meno significative.

Basta anche solo la tabella dei risultati (dove si parte e passaggi intermedi non è così rilevante).

Formule del modello in R:

* Fit\_lm <- lm(…) si possono riportare sopra il risultato di R.

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)

Riferimenti:

* Al repository github (?)
* Citare il libro da cui sono stati presi i metodi (libro di apprendimento statistico)

COMMENTI PARTICOLARI AL REPORT

Tabella introduzione:

* Apici
* Nome colonna
* Allineamento sinistra scritte e destra se sono numeri
* Termine feature ok in statistical learning
* Prima descrizione e poi acronimo
* Nome regressore non va scritto senza dire cosa è -> games played (“GP”)

Il nome della variabile va indicato almeno una volta nel testo con l’acronimo tra parentesi per poi usare solo l’acronimo

Sezione 2:

* Columns selection, non cancellation

Sezione 3:

* Sezione linear regression
* Ingrandire immagine e allungare caption dell’immagine della correlazione
* Istogramma delle feature importante
* Screenshots con sfondo nero -> diventare tabelle (cifre decimali)
* A, b, c in subplot vanno descritti nella caption della figura (esempio: figura 4)

A, b, c da sole sotto l’immagine e la descrizione va riportata sotto.

* Model = equazione -> cambiare

Sezione 4:

* Bar plot non equamente distribuite: non un problema
* Figura 13: più esplicativo

Probabilità di previsione da spiegare perché sono mostrate.

Vedere le probabilità di previsione per le 2 classi per vedere se sono ben separati e mostrare con degli istogrammi per maggiore chiarezza.

* Istogrammi di probabilità per 2 classi -> probabilità che si compensano
* Separare istogrammi per le 2 variabili.

Sezione finale:

* Provare svm

Va bene quello che c’è, non è ripetitivo.

TODO

* Sistemare nomi delle tabelle
* ~~Scientific questions nell’introduzione~~
* ~~Positive variables nell’introduzione~~
* 2 cifre decimali ovunque con “e-…”
* Variable e non coefficient nei nomi delle tabelle
* Sistemare ref a table nel codice
* Backward cambia qualcosa?
* Figure -> titolo, axis labels, controllare font
* Istogrammi più centrati
* Alberi: GP descrivere perché split su .5
* Appendice modelli iniziali
* Variazioni % R^2 e MissErrRate
* Riferimenti libro e dataset
* Rispondere nella conclusione

SEZIONE 4 COMMENTI DEL NOSTRO REPORT !!!!!!!!!!!!!