PTDL - Đỗ Như Tài Phương pháp chấm:

⁃ Giữa kỳ

◦ Thu thập dữ liệu và hiểu dữ liệu (EDA)

⁃ Cuối kỳ

◦ Tiền xử lý và mô hình máy học

=> Cốt lõi của môn này tập trung vào làm thế nào để hiểu dữ liệu và kể câu chuyện từ dữ liệu, còn mô hình thì chỉ cần ứng dụng hợp lý phù hợp nhất.

⁃ CAPM và mật độ ảnh hưởng của tài sản Beta so với thị trường

⁃ Hiểu về variance (population và sample)

⁃ Dựa vào variance để phân tích lợi nhuận và rủi ro. Từ đó dựa vào để ra được (Sharpe Ratio)

=> Lợi nhuận / rủi ro

⁃ Rút ra được trường hợp rủi ro bằng 0 (profit => lợi nhuận) nếu xét về trục hoành thì gần sát

⁃ So sánh khái niệm phân tán và phương sai có ví dụ thực tế CÁC DẠNG PHÂN TÍCH

⁃ Hồ sơ, pháp lý: cần đưa ra các feature về nhân quả

⁃ Market, sale: cần đưa ra feature về correlation và so sánh

=> Việc sử dụng KPI card giúp hiểu được mối liên hệ của chúng từ đó nhìn từ con số để đưa ra kết quả

NHỮNG VẤN ĐỀ KHÁC

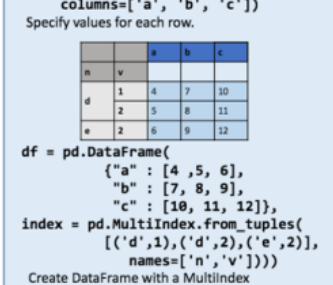
⁃ hồi quy tuyến tính, ….

⁃ Ví dụ về các tính một f nào đó trong đó biết về x và y rồi

⁃ Đối với hình ảnh, văn bản, thời gian xử lý như thế nào

LÀM BOOK (1-12) CHỪA CHƯƠNG CUỐI (BÀI TẬP) VÀ LÀM DUY NHẤT MỘT FILE IPYNB CÓ CODE GÕ HẾT VÀ TẠO MỘT FILE HTML





* Phân tích bài toán Iris
* Làm file word để nói lý thuyết nào => minh chứng đâu
* Các cột và dữ liệu đó là gì (head(5) head(-5)

EDA

* Đặt câu hỏi
* Trả lời câu hỏi dựa vào mô hình/ biến đổi
* Đặt ra câu hỏi mới
* ………..
* Xây dựng câu chuyện về data (PP LUẬN)

CAC KIEU SO SANH

* 4

GHI AM (PAGE23)

THIEEUS:

* SO SANH CAC LOAI (CLASSIFICATION) XEM GIA TRI CUA NOS NTN
* Ve chart scatter about cac moi features vs target
* Cac cot nay tinh the nao?
* Giai thich cac cot nay la gi?
* Ve hinh moi noi ra
* Nguyen nhan ly do cac cot co gia tri 0 bat thuong
* Ve them chart so sanh bang bieu do phan chia cua cac features vs target (histogram)

(ty le giua 1/0) Y: target/ x:feature => tim nguong (ro hon) => chia nguong de danh gia

* Xet them cac bien co anh huong cao => tim hieu them la cac bien do co su gia tang dua vao bien nao khong de tang xac suat
* Doi voi insulin la xet them cac yeu to khac (vd: insulin cao van bi?) yeu to nao va tai sao tang giam
* Tim hieu them ve cac chi so do do moi tuong quan

WORD:

TItLE above Table

Image above Title

+

DU LIEU SO => CHIA NGUONG

**Data Preparation and Feature Manipulation Techniques**

Fixed\_acidity (FA)

**Tương quan dương**

* **citric\_acid (0.67):** Citric là thành phần của fixed acidity.
* **density (0.67):** Nhiều chất hòa tan (axit, muối, đường…) tăng, mật độ tăng. Lô rượu có FA cao thường có tổng chất hòa tan cao hơn.
* **residual\_sugar (0.11):** Đường dư tăng làm density tăng; mà density có xu hướng dương với FA
* **Sulphates (0.18):** Một số nhà làm rượu sẽ có xu hướng giữ trong mẻ rượu có nhiều acid cố định (tươi, bền, an toàn vi sinh) vì thế nên tổng chất hòa tan (ions, muối) cao. Mà sulphates lại là muối trong rượu, cho nên có độ tương quan nhẹ.

**Tương quan âm**

* **pH (-0.68):** FA tăng thì lượng axit trong rượu cao, mà lượng axit cao thì pH giảm.
* **free\_SO2, total\_SO2 (-0.15, -0.11):** Hiệu lực diệt khuẩn SO2 tùy thuộc vào pH, khi pH cao thì hiệu lực diệt khuẩn S02 giảm. Để giải quyết vấn đề này các nhà làm rượu thường bổ sung nhiều SO2 khi pH cao, mà pH cao thì acid thấp và ngược lại.
* **volatile\_acidity (-0.26):** chủ yếu là acetic acid sinh ra từ nhiễm/oxy hóa. Để ức chế vi khuẩn acetic người ta cần môi trường pH thấp, mà pH thấp thì lượng fixed acidity cao.