


ML-2 Interview Prept Practical Session [Start at 9:10 pm]

⇒ ↗ Recommendation ↗ TS Forecasting ↗ ↘

Q Case Study → Q How will you design a Rec. System
Q How will — || — Forecasting model *

→
⇒ Time series

Q Explain ARIMA → $\hat{P}, \hat{Q}, \hat{d}$, Stationarity

Q Assignment Question

Interview Process

⇒ Junior level → • Classical ML → Regression
• very basic ARIMA

⇒ Senior level → Designing Forecasting Models:

↳ Are around Scale: -
→ Regression, SARIMAX, Residuals → 210 Rev. → 1B

Q Design a Model for a online/offline retail store
↳ Clothing :-

Q Store data → what datasets would be used in TS
• and how
→ Size → exogenous variable



Q What is parquet?

- Zipped data
- Column based data storage
- Fast to read

Size of your data:

- Inputs will your model take

Q Design a model →

- ↳ overall Sales
- ↳ Store level Sales
- ↳ Store + Product level Sales

⇒ Store + Category level

⇒ Sparcity :- a lot of blank data.

→ Limited edition	<u>Shoes by adidas</u>	→ 15,000 no. of sales	Part P
→ S1	I2	15000	3
			18 May
		4	19 May
		2	20 May
			21 May
			=

→ Shoes by adidas ↗ at higher level

↳ This is better

Chance of

data being flwr,

30 18 May
40 19 May
10 20 May
20 21 May

⇒ Q: Why sparsity is a problem in Forecasting ??

↳ TS forecast → works only with continuous TS.

Q: What are the pre processing steps for Forecasting ??

↳ Imputation

↳ clip for outlier

↳ Convert to datetime

→ X → Y →

\Rightarrow Store S1	Product P1	\rightarrow 18, 19, 28 May
S2	P1	\rightarrow 21, 22 May
S1	P2	\rightarrow 01 June

\rightarrow fill missing dates in the dataframes

Creating dates in my df

\Rightarrow 2020-10-01	200
-02	300
-04	400

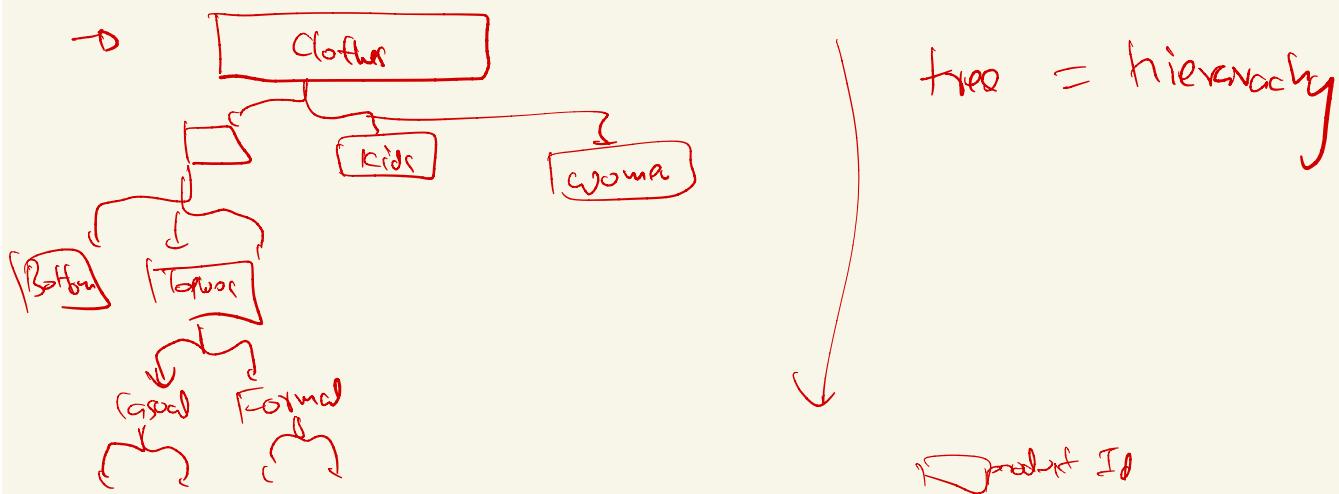
2020-10-01	200
-02	300
-03	0
-04	400

Q Show you a TS plot

\hookrightarrow What do you see!!

- Trend \rightarrow
- Seasonality \rightarrow
- Missing \rightarrow
- Outliers \rightarrow

\Rightarrow Hierarchy \rightarrow Structure of a item range.

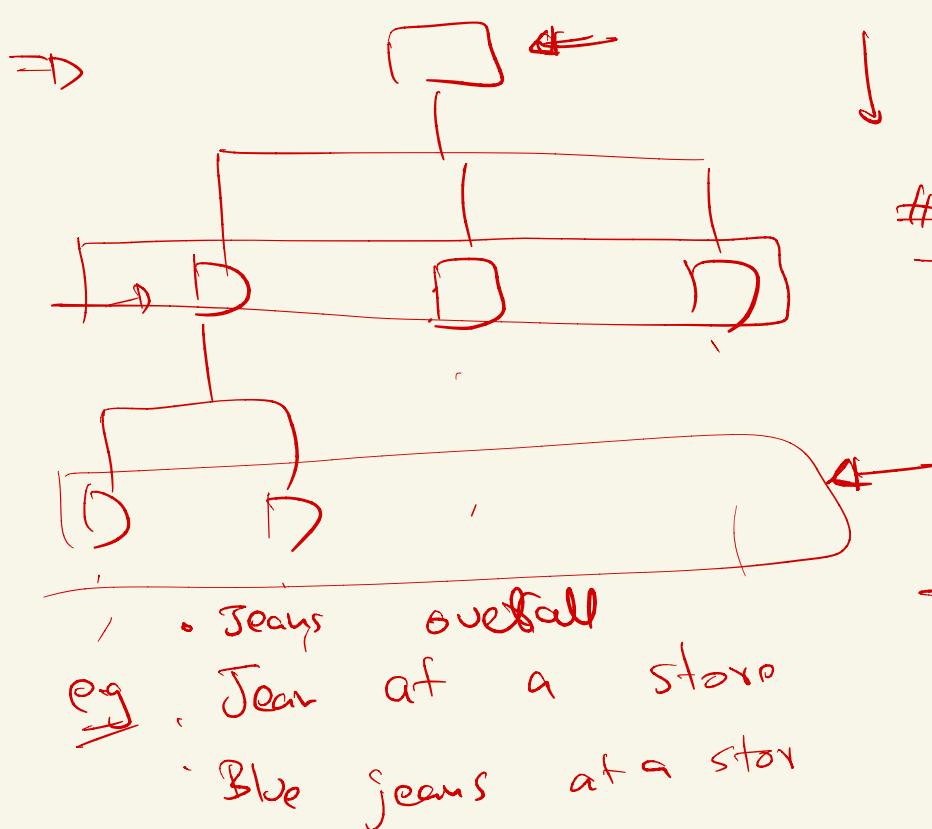


Q How much ram is needed for a 250gb stored data in memory =

$$\text{Typical} \quad 2x \quad = \quad 500 \text{ gb} \quad [\text{RAM}]$$

Q Reduce Columns \rightarrow Reduce memory \rightarrow Inverse / Higher Rank

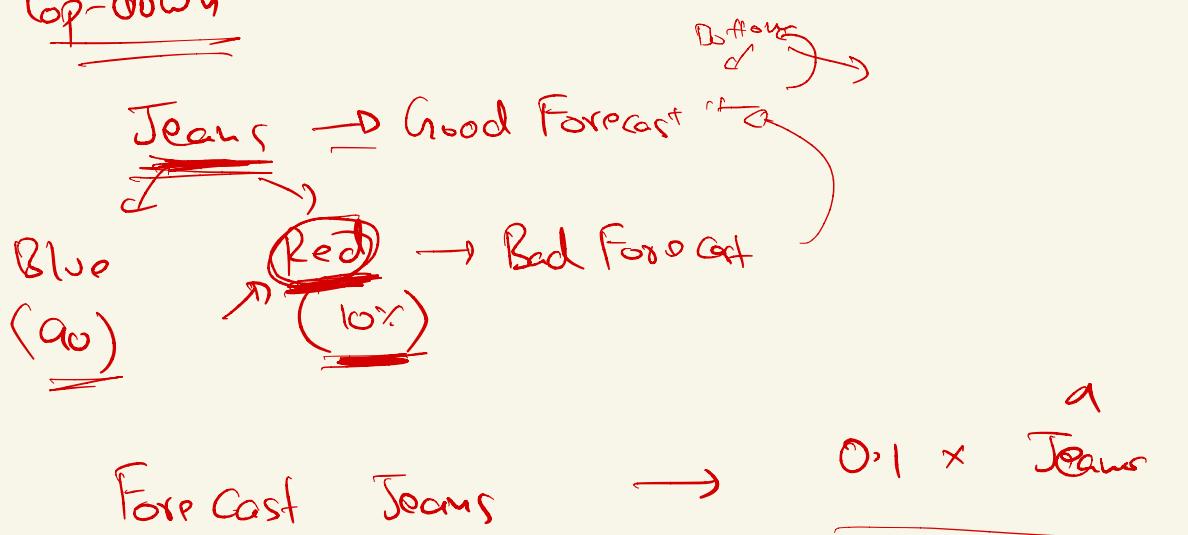
\Rightarrow Hierarchical Forecasting



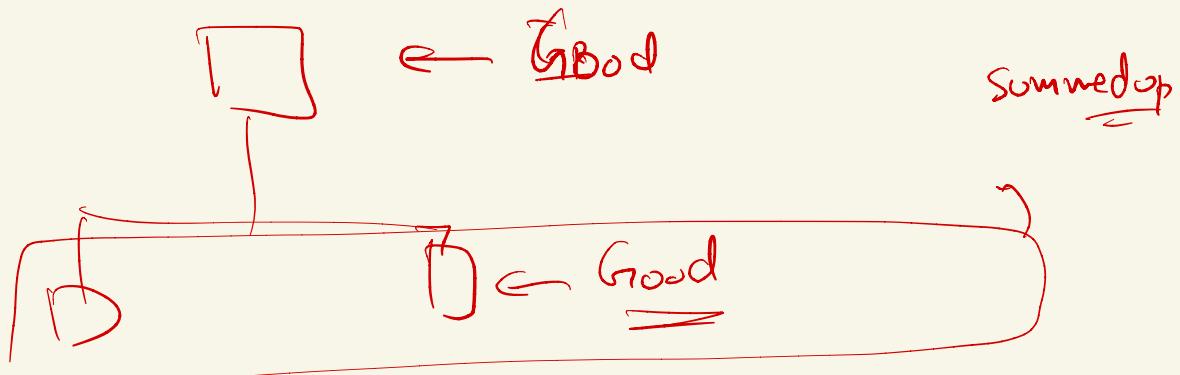
as we go down
↓
• Sparsity ↑
 $\frac{\# \text{ of TS}}{\# \text{ of Datapt / series}} \uparrow$

\rightarrow as go down
↓
its hard to predict

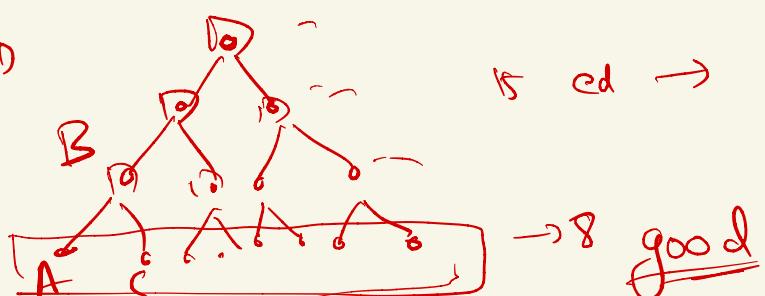
⇒ ① Top-down



② Bottoms up hierarchy



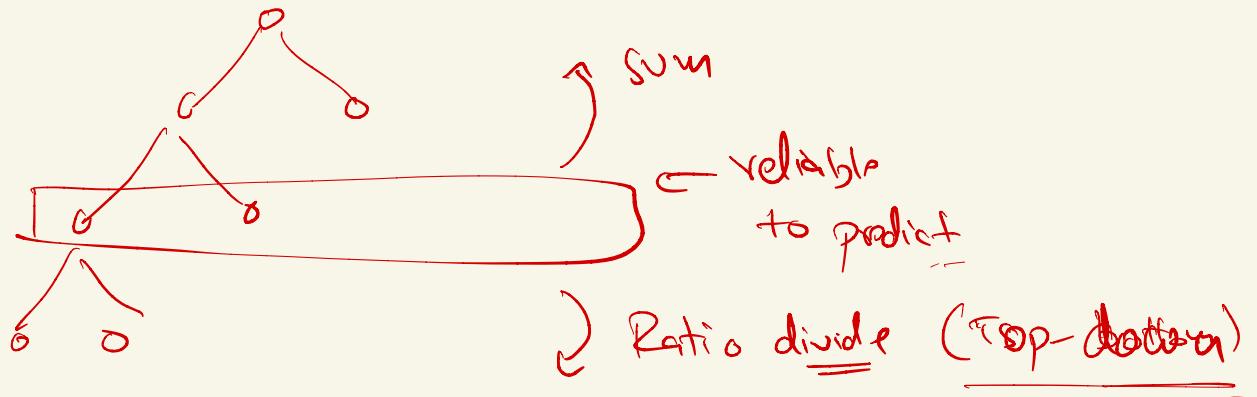
⇒ IS cd → IS mde



$$\text{Sum} \rightarrow \underline{A + C = B}$$

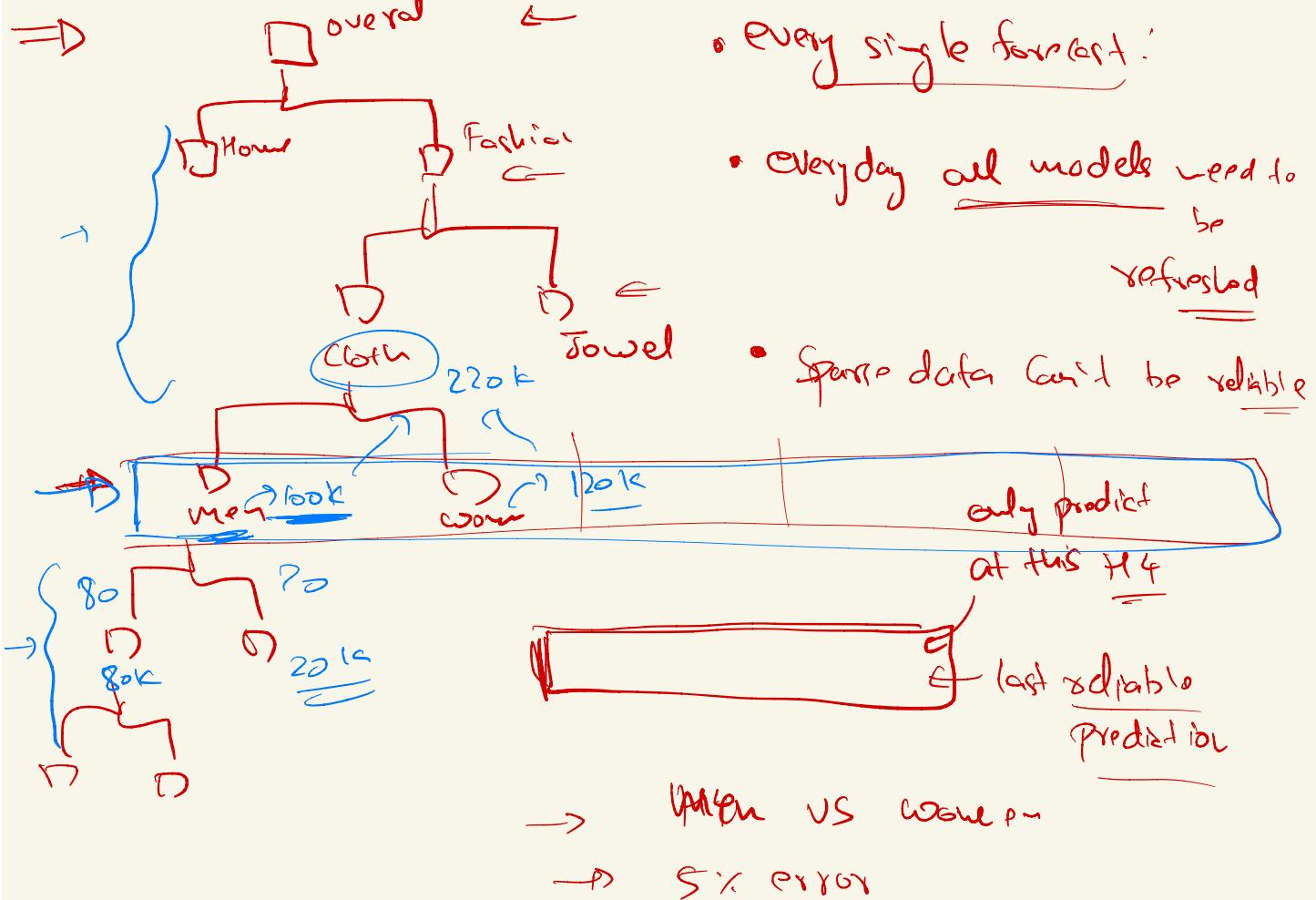
\Rightarrow Middle Out \rightarrow

$\underline{H4} \rightarrow \underline{5\% \text{ error}}$



\Rightarrow overall

• Every single forecast!

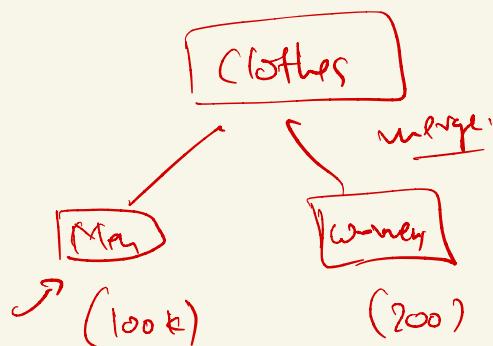


Q How do you forecast Sparse TS in an ecommerce store?

↳ Predict at a higher hierarchy

↓
Ratio to find actual forecast.

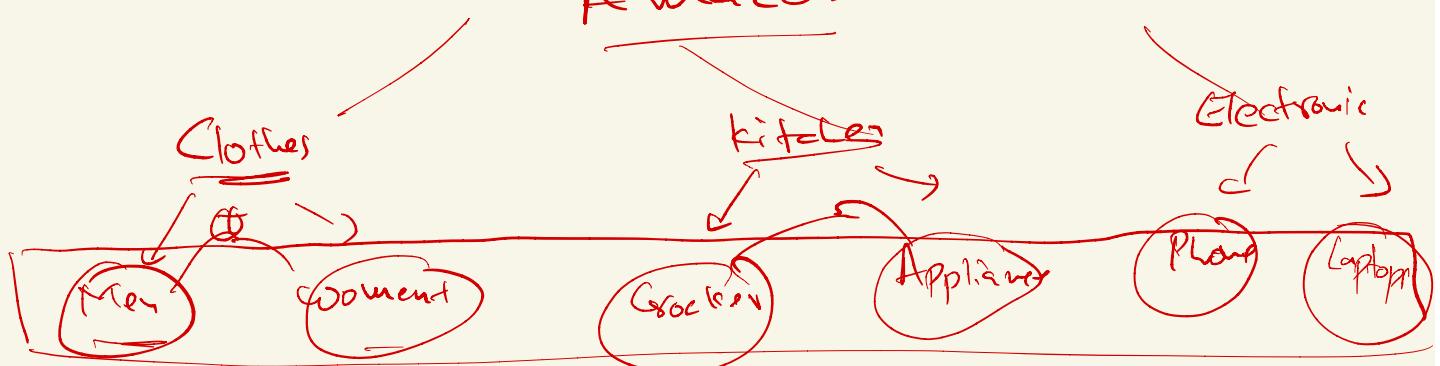
⇒ middle → Top



$$100 + 200$$

predicting at level → all possible forecast at first level

Amazon



Home
Work



- TES *
- SARIMAX *
- Prophet