Challenges in Machine Learning | Problems in Machine Learning

10 important Point

1)Data Collections

Csv files, fetch data through API, or fetch data through web scrapping

2)Insufficient Data/labelled Data

3)Non-Representative Data

Is main data adhi kahani bolta hai

Isko sampling Noise bhi kehtay hain

Example yeh hai iska kay agar hum survey karain kay 2026 ka t20 world cup kon jeta ga or hum sirf survey Pakistan main tu Zahiri si bath kay most people paksitan ko support karaingay usolan humain har us country ian karna chaiye jahan teams match khailty hain

4)Poor quality Data

Data ki quality ka sahi hona bohat zarori hai

If the quality of data is poor our result will be poors

5)Irrelevant features(columns)

Aisay columns jinki humain zarorat hi nhi hoti unks koi requirement ya contribution hi nhi hoti islie hum unhain remove kartay hain

“Garbage In Garbage Out” Model kay andher khachra dalogay tu khacra hi miliga

6) 🟥 **1) Overfitting — (Model ne data ko “yaad” kar liya)**

Roman Urdu:

Overfitting tab hota hai jab model **training data ko hadd se zyada seekh leta hai**, hatta ke **noise, mistakes, useless details** bhi yaad kar leta hai.

Nateeja:

- Training data per **accuracy high**
- New data (test data) per **accuracy low**

➡️ Yani model **exam ke past papers ratta** kar leta hai, concept nahi samajhta.

Example:

- Cat ki image pe model ne whiskers ka exact shape yaad kar liya.
- Nayi image me whiskers thori alag hojae to model confuse.

7)Underfitting — (Model ne data ko "seekha hi nahi")

Roman Urdu:

Underfitting tab hota hai jab model **itna simple hota hai** kay woh training data ki **real patterns hi seekh nahi pata**.

Nateeja:

- Training data per **poor performance**
- Test data per bhi **poor performance**

➡️ Yani model ne "ratta bhi nahi lagaya aur concept bhi nahi seekha".

Example:

- Tum linear line laga rahe ho jab data curved shape mein ho.
- To line sahi fit hi nahi hoti.

- **Concept Overfitting Underfitting**

Learning Learns too much (even noise) Learns too little

Model	Too complex	Too simple
Training Accuracy	High	Low

Low Test Accuracy Low

Reason Extra patterns memorized Real patterns not learned

8)Software integration

9)offline learning /deployment

challeng

10)cost involved