# Predictive Analysis

## UNIT-2

### Prediction effect:

=> Predictive models influence real world outcomes by enhancing efficiency, reducing risks, profitability and decision making when deployed.

Retail Ex: 1) Forecasts customer demand to optimize stock and minimize waste

2) Healthcare identifies at risk patients early for better preventive care

3) Fraud detection identifies suspicious patterns to prevent losses.

Key measures:

* Prediction accuracy
* Customer satisfaction
* Operational cost reductions
* Business growth.

Additional Benefits:

=> Improves resource allocation, enhances customer personalization, supports proactive ~~strategy~~ strategies.

---

### Deployment of Prediction model:

Overview: Integrates trained model into operations for real-time or batch predictions.

=> Put trained model into daily use for quick predictions.

Key steps:

1) Validate on new/unseen data.

2) Test scalability for big data volumes.

3) Embed in apps, dashboards or systems.

4) Monitor performance drifts and retain

5) Ensure security and compliance during rollout

Real Ex: Bank auto-approves loans with credit score

Challenges:

* Handling data drift
* Version Control
* User training
* Data changes
* Staff training

---

Ethical Concerns in P-A:

=> PA involves using data and models to forecast outcomes, but raises ethical issues around privacy bias, transparency and accountability

Due to data handling and decision impacts

Key Concerns:

1) Privacy and surveillance: Models use personal data like health records or location without full consent, risking

re-identification even if anonymized, leads to intrusive monitoring.

2) Bias and Discrimination: Training on biased historical data amplifies unfairness.
Ex: loan denials or hiring rejections for certain groups.

3) Black Box Secret: Models decide things you can't understand or explain, like why a loan was denied

4) Accountability Gaps: Unclear who is responsible for harmful predictions.
Eg: wrongful arrests from flawed models.

5) Misuse of Prediction: Can exploit vulnerable people or prioritize company profits over well-being.

Ethical Responsibilities:

1) Ensure Data Privacy and Security: Lock data tight, follow laws like GDPR to protect info.

2) Obtain informed Consent: Explain exactly what data is for and get clear user permission.

3) Mitigate Bias and Promote Fairness: Use balanced data, test models to treat everyone equally.

4) Promote transparency: Create models that explain decision clearly Eg: "why this choice".

5) Establish Accountability: Set rules on who's responsible and how to handle mistakes.

6) Regularly Audit and Update Models: Check models often, fix issues, keep them fair.

7) Assemble a Diverse Team: Hire people from different backgrounds to catch all problems.

---

The Data Effect:

⇒ Means how data quality and features directly shape predictive model success good data gives accurate results, bad data causes errors and bias.

Important aspects of Data Effect:

1) Data Quality: clean, complete, consistent data improves accuracy.

2) Data Diversity: Covering multiple scenarios avoid bias.

3) Data preprocessing: Includes cleaning, normalization, handling missing values and removing noise

How the Data Effect Influences Predictive Models:

1) Class Imbalance: Rare cases underrepresented, models bias toward majority class, poor minority predictions.

2) Data Quality: Inaccuracies, incompleteness, inconsistencies cause wrong or unreliable results.

3) Data Privacy: Sensitive data restrictions limit training data amount and type.

4) Data volume: Large data hard to process/ store; small data misses patterns.

5) Data variety & Velocity: Diverse types structured/ unstructured and fast streams add complexity.

Practices to Mitigate Related Issues:

For Class Imbalance:

1) Oversampling/ Undersampling: SMOTE (Synthetic Minority Over-sampling Technique) creates synthetic minority data.

2) Cost-sensitive Learning: Higher penalties for minority errors.

For Data Quality Issues:

1) Data validation and Cleaning: Fix errors, missing values.

2) Feature Engineering: Build Better features from raw data.

For Data Privacy Concerns:

1) Data Anonymization: Remove identifiers via aggregration.

2) Differential Privacy: Add noise for privacy protection.

3) Secure Multi-Party Computation: Share insights without revealing data

General Practice:

1) Domain Expertise: Experts guide data understanding.

2) Model Validation: Cross-validation on new data.