

Introduction to Predictive Analytics

Introduction to Predictive Analytics

Key concepts, workflow, tools, and real-world applications

1. What Is Predictive Analytics?

Aspect	Description
Definition	The use of statistical algorithms, machine learning models, and data mining techniques to forecast future outcomes based on historical and current data.
Goal	Turn data into actionable insights that anticipate events, behaviors, or trends.
Difference from Descriptive & Prescriptive Analytics	- Descriptive: "What happened?" (e.g., dashboards, reports). - Predictive: "What is likely to happen?" (forecasting, scoring). - Prescriptive: "What should we do?" (optimization, recommendation engines).

2. Core Components

1. Data

- Structured (databases, spreadsheets).
- Unstructured (text, images, sensor logs).
- External sources (social media, market data).

2. Feature Engineering

- Cleaning & preprocessing (missing values, outliers).
- Transformation (normalization, encoding).
- Creation of new variables (lag features, aggregates).

3. Modeling Techniques

- **Statistical:** Linear regression, logistic regression, ARIMA.

Introduction to Predictive Analytics

- **Machine Learning:** Decision trees, random forests, gradient boosting, support vector machines, neural networks.
- **Time Series Specific:** Prophet, LSTM, SARIMA.

4. Evaluation Metrics

- **Regression:** RMSE, MAE, R^2 .
- **Classification:** Accuracy, Precision, Recall, F1 score, AUC/ROC.
- **Ranking/Scoring:** MAP, NDCG.

5. Deployment & Monitoring

- Model serving (REST API, batch scoring).
- Performance drift detection.
- Retraining pipelines.

3. Typical Predictive Analytics Workflow

1. Problem Definition

- Business objective ? measurable KPI.
- Determine prediction horizon & granularity.

2. Data Acquisition

- Identify internal & external data sources.
- Set up ETL pipelines (extract, transform, load).

3. Exploratory Data Analysis (EDA)

- Summary statistics, visualizations, correlation analysis.
- Spot patterns, seasonality, anomalies.

4. Data Preparation

- Handle missing values (imputation, deletion).
- Encode categorical variables (one-hot, target encoding).
- Scale/normalize numeric features.

Introduction to Predictive Analytics

5. Feature Engineering

- Domain-specific transformations.
- Temporal features (day-of-week, holidays).
- Interaction terms, polynomial features.

6. Model Selection & Training

- Baseline model (e.g., simple linear regression).
- Hyperparameter tuning (grid search, random search, Bayesian optimization).
- Cross-validation to assess generalization.

7. Model Evaluation

- Compute appropriate metrics on validation set.
- Compare models, check for overfitting/underfitting.

8. Interpretability & Validation

- Feature importance (SHAP, LIME).
- Business validation: does the model make sense to stakeholders?

9. Deployment

- Export model (pickle, ONNX, PMML).
- Integrate into production system (microservice, batch job).

10. Monitoring & Maintenance

- Track prediction accuracy over time.
- Alert on data drift or performance decay.