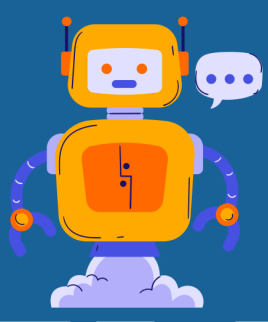


# MACHINE LEARNING

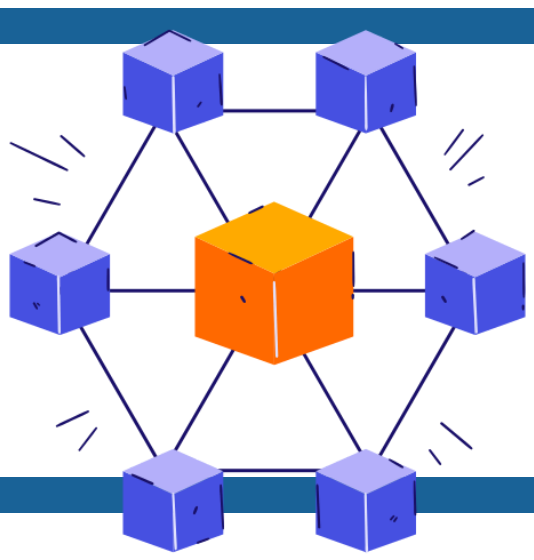
## PIPEPLINE

### DATA COLLECTION

- Gather raw data from various sources (databases, APIs, files, sensors)
- Ensure data quality and completeness
- Handle data versioning and lineage

### DATA PREPROCESSING

- Data Cleaning: Handle missing values, outliers, duplicates
- Data Transformation: Normalization, standardization, encoding
- Feature Engineering: Create new features, select relevant ones
- Data Splitting: Train/validation/test sets



### TRAINING

- Algorithm selection and hyperparameter configuration
- Cross-validation for robust evaluation
- Model fitting on training data

### EVALUATION

- Performance metrics (accuracy, precision, recall, F1, AUC, etc.)
- Validation on unseen data
- Model comparison and selection

### DEPLOYMENT

- Performance metrics (accuracy, precision, recall, F1, AUC, etc.)
- Validation on unseen data
- Model comparison and selection