2025

Team 3

AI-Driven Cost Estimation for Injection Mold Quotation

AGENDA

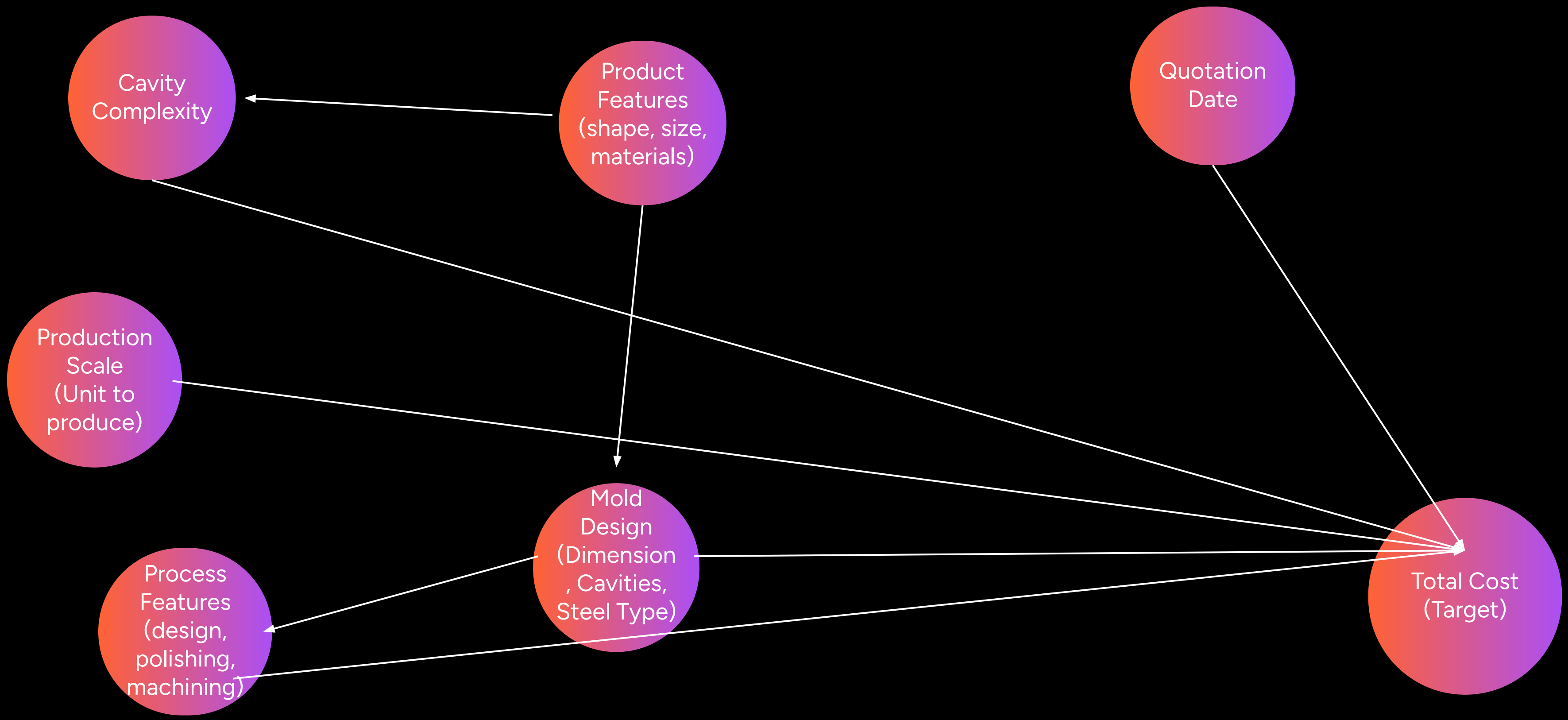
01 Model, Data, Algorithm, Workflow

02 Prediction Precision

03 Impact Study (Environmental, Cyber risk, change management, a control and maintenance plan

04 Deliverables: Variable dictionary, Data quality diagnosis, training survey, performance analysis, impact study

Theoretical Causal Model



Theoretical Causal Model

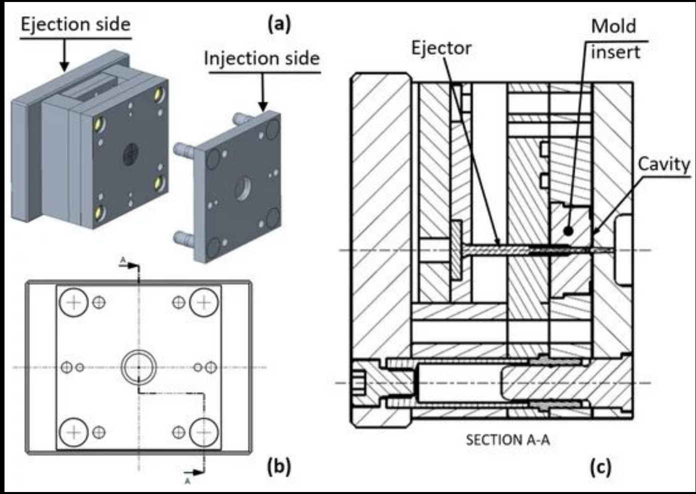
1. Product Features (shape, size, material)
→ Influence **Mold Design** (dimensions, cavities, steel type)
→ Affect **Cavity Complexity** (e.g. intricate geometries increase complexity)

2. Mold Design
→ Directly contributes to **Total Cost** (due to size, number of cavities, steel type)

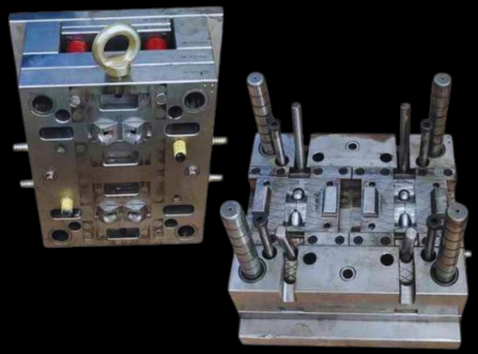
→ Influences **Process Features** required (design effort, polishing, machining)

3. Process Features (design, polishing, machining)

→ Directly impact **Total Cost** (higher process requirements = higher cost)



Theoretical Causal Model



4. Production Scale (units to produce)
→ Directly impacts **Total Cost** (due to economies of scale or complexity adjustments)

5. Cavity Complexity

→ Direct effect on **Total Cost** (complex cavities require more work and precision)

6. Quotation Date

→ Impacts **Total Cost** (price fluctuations over time, inflation, material market variations)

It might be considered as a contextual variable.

Data

What we did:

- Selected key features from 70+ columns
- Handled missing values
- Encoded categorical data (like steel and plastic types)
- Transformed `Total_price` using log scale
- Split data into training and test sets

What was tricky:

- Some features were unclear or redundant
- Many binary flags (0/1) with little explanation
- Hard to know what really impacts cost

Algorithms

Model Results & Insights

Approach:

- Built regression models using historical quotation data (n = 643).
- Tried both **Linear Regression** and **Random Forest Regressor**.

Results:

- **Linear Regression R^2**: ~0.21
- **Random Forest R^2**: ~0.33

WE ARE STUCK HERE

Algorithms

Model Results &

Approach:

- Built regressi
  (n = 643).
- Tried both **Li**
  Regressor.

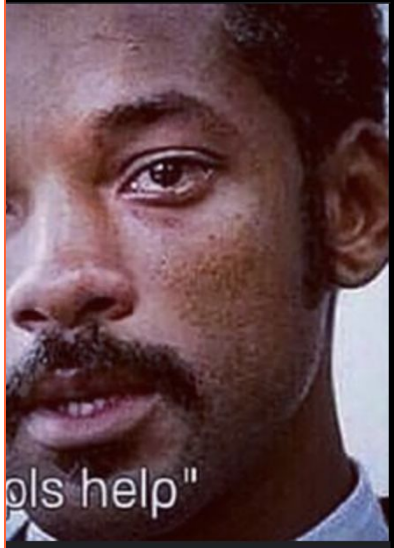
Results:

- **Linear Regre**
- **Random For**

WE HAVE PROGRESS

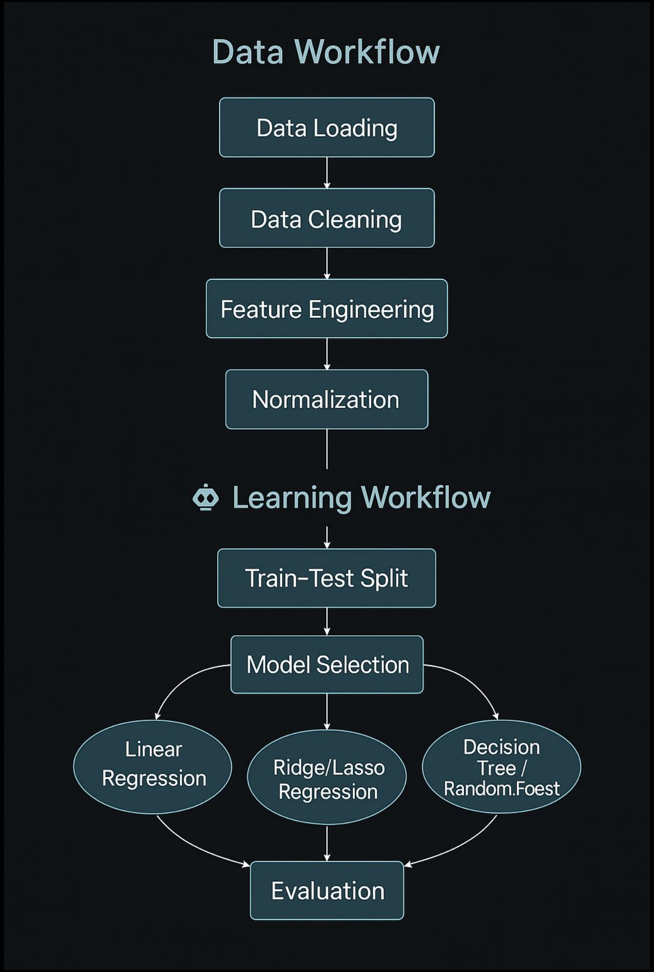
WE'RE STILL TRYING

WE NEED MORE TIME



WE ARE STUCK HERE

Workflow


Data Workflow
Data Loading
Data Cleaning
Feature Engineering
Normalization
Learning Workflow
Train-Test Split
Model Selection
Linear Regression
Ridge/Lasso Regression
Decision Tree / Random.Foest
Evaluation

Workflow

1. Data Loading

- **We use** Python with **pandas** (`pd.read_csv`)
  Good formatting for Jupyter.

2. Data Cleaning

- Handle **missing values** (NA and ZEROS)

3. Feature Engineering

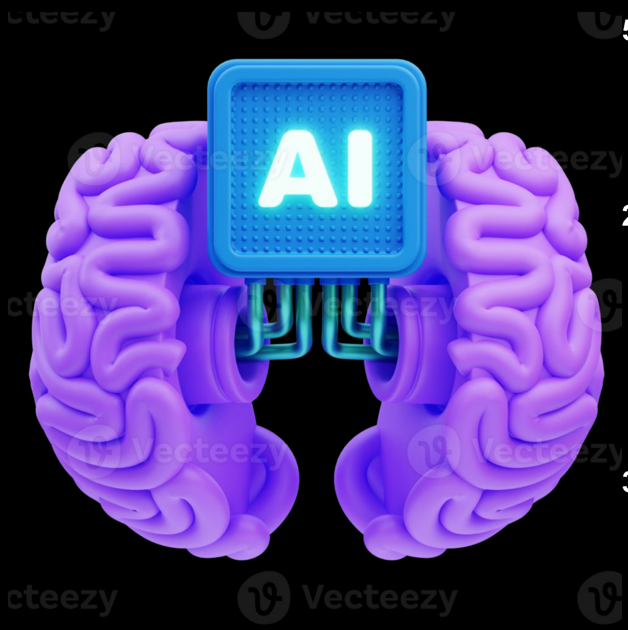
  - **We** Extract the necessary variables for the modeling that affects the price.

    WE SELECTED: `'ACHATS', 'AJUSTAGE', 'CN', 'ETUDES', 'METHODES', 'MOULE.CLEF.EN.MAIN1', 'ST', 'Mold_Volume_A4'`

4. Scaling process

- Scale numerical features to **mean=0, std=1** (StandardScaler) with a function in Sklearn.

Workflow



5. Splitting

- We Create **train-test splits** using scikit-learn's `train_test_split`, ensuring temporal splits if modeling time-sensitive prices.

2. Model Selection

- **Candidates:** Linear Regression and Random Forest Regressor
- **Tool:** scikit-learn (`sklearn.linear_model`, `sklearn.ensemble`)
- **Rationale:** Start simple (Linear)
- And Increase complexity (Tree-based) for Random Forest

3. Evaluation & Impact

→ Compare performance.

Impact Study (Environmental Risk)

- Mold production uses steel, energy, and water.
- AI-driven quoting may influence environmental outcomes indirectly.

Positive Impacts

Reduces overproduction and material waste.

May favor simpler, more sustainable designs.

Can evolve to include CO_2 or energy metrics.

Risks

Current model ignores sustainability factors.

May favor cheapest options, not greenest.

No data yet on energy or material footprint.

Future Steps

Add eco-metrics (energy, scrap rate).

Apply Green AI practices.

Explore sustainability-focused extensions.

Impact Study (Cyber Risk)

- The system handles sensitive industrial data (designs, prices, materials).
- Risk increases when deploying AI tools online or across teams.

Benefits	Risks	Mitigation Strategies
Reduces overproduction and material waste.	Exposure of proprietary cost models or client quotes.	Implement access control and encryption.
May favor simpler, more sustainable designs.	Risk of unauthorized access to sensitive product specs.	Regular security audits and version tracking.
Can evolve to include CO_2 or energy metrics.	Possible model manipulation or data poisoning.	Log user interactions for traceability.

Impact Study (Change Management)

- Switching from expert-based to AI-driven quoting affects workflows and roles.
- Resistance may come from mold designers, sales, or finance teams.

Benefits	Risks	Recommendation
Standardizes quoting process across teams.	Risk of mistrust or rejection by experienced staff.	Involve experts early in system design.
Reduces reliance on individual expertise.	May reduce flexibility in custom quoting.	Provide training and transparent model explanations.
Speeds up decision-making with consistent logic.	Requires retraining and mindset shift.	Allow manual overrides or hybrid decision options.

Impact Study (Control & Maintenance Plan)

- AI models degrade over time (data drift, outdated assumptions).
- Regular updates ensure accuracy and trust.

Key Actions

Schedule periodic model retraining (e.g., every 6–12 months).

Monitor performance metrics continuously.

Log predictions and feedback to detect errors.

Data & Model Management

Maintain a version-controlled dataset and model registry.

Track changes in features (e.g., new steel types or processes).

Ensure compatibility with evolving business needs.

Governance

Assign ownership (data steward, model owner).

Define escalation paths for anomalies or failures.

Include experts in update reviews.

Deliverables

Variable dictionary, Da dy

Work in Progress