flowchart TD

%% Layer 1 - Default Processing

A[Raw CSV File] --> B[Layer 1 - Default Pipeline]

B --> C[Detect Encoding]

C --> D[Identify Delimiter]

D --> E[Validate Headers]

E --> F[Handle Multiline Cells]

F --> G[Remove Extra Whitespace]

G --> H[Load into DataFrame]

%% Column Classification

H --> I{Column Classification}

%% Text Processing Branch

I --> J[Text Columns]

J --> K[Convert to String]

K --> L[Lowercase Strip]

L --> M[Remove HTML Tags]

M --> N[Unicode Normalization]

%% Other Column Types

I --> O[Numerical Columns]

I --> P[DateTime Columns]

I --> Q[Categorical Columns]

%% Preserve as Metadata

O --> R[Preserve as Metadata]

P --> R

Q --> R

%% Layer 2 - User Enabled

R --> S[Layer 2 - User Enabled]

N --> T[User Options]

%% User Options Processing

S --> T[User Options]

T --> U[Preprocess Text Columns]

U --> V[Text Normalization]

V --> W[Stemming]

W --> X[Lemmatization]

X --> Y[Stop Word Removal]

T --> Z[Data Type Validation]

T --> AA[Duplicate Row Deduction]

T --> BB[Missing Values Handling]

BB --> CC{Missing Value Strategy}

CC --> DD[Leave As Is]

CC --> EE[Fill Custom Values]

CC --> FF[Drop Rows or Columns]

%% Metadata Collection

DD --> GG[Metadata Collection]

EE --> GG

FF --> GG

%% Quality Gates

R --> HH[Quality Gates]

Y --> HH

Z --> HH

AA --> HH

GG --> HH

HH --> II[User Review Point]

%% Quality Decision

II --> JJ{Quality Pass?}

%% Success Path

JJ -->|Pass| KK[Final Clean DataFrame with Metadata]

KK --> LL[Proceed to Chunking Layer]

%% Failure Path

JJ -->|Fail| MM[Error Handling and User Notification]

MM --> NN[User Action]

NN --> OO{Action Type}

OO -->|Retry| S

OO -->|Adjust Config| PP[Update Configuration]

PP --> S

%% Styling

classDef layer1 fill:#e1f5fe,stroke:#01579b,stroke-width:2px

classDef layer2 fill:#f3e5f5,stroke:#4a148c,stroke-width:2px

classDef textProcess fill:#e8f5e8,stroke:#1b5e20,stroke-width:2px

classDef decision fill:#fff3e0,stroke:#e65100,stroke-width:2px

classDef metadata fill:#fce4ec,stroke:#880e4f,stroke-width:2px

classDef quality fill:#fff8e1,stroke:#f57f17,stroke-width:2px

classDef error fill:#ffebee,stroke:#b71c1c,stroke-width:2px

classDef success fill:#e0f2f1,stroke:#004d40,stroke-width:2px

class A,B,C,D,E,F,G,H layer1

class S,T,U,V,W,X,Y,Z,AA,BB layer2

class J,K,L,M,N,V,W,X,Y textProcess

class I,CC,JJ,OO decision

class R,GG metadata

class HH,II quality

class MM,NN,PP error

class KK,LL success