

```
import polars as pl
from sklearn.model_selection import train_test_split
from sklearn.impute import MissingIndicator
from recipes import Ingredients, Recipe
from recipes.selector import has_role
from recipes.step import StepImputeFill, StepScale, StepSklearn

# Load and split Physionet Computing in Cardiology Challenge 2019 dataset
df = pl.read_csv("Physionet_CiCC_2019.csv", sep="|")
df_train, df_test = train_test_split(df, test_size=0.2, random_state=42)

# Define ingredients (training set + roles)
roles = {"outcomes": ["SepsisLabel"], "predictors": ["Age", "HR", "Temp"], "groups": ["PatientID"],
"sequences": ["ICULOS"]}
ing = Ingredients(df_train, roles=roles)

# Define the recipe for processing the ingredients (e.g., scaling and imputation)
rec = Recipe(ing)
rec.add_step(StepScale())
rec.add_step(StepSklearn(MissingIndicator(features="all"), sel=has_role("predictor")))
rec.add_step(StepImputeFill(strategy="forward"), sel=has_role("predictor"))

# Prepare (=fit) recipe on training data and apply the same recipe to the test set
df_train = rec.prep()
df_test = rec.bake(df_test)
```