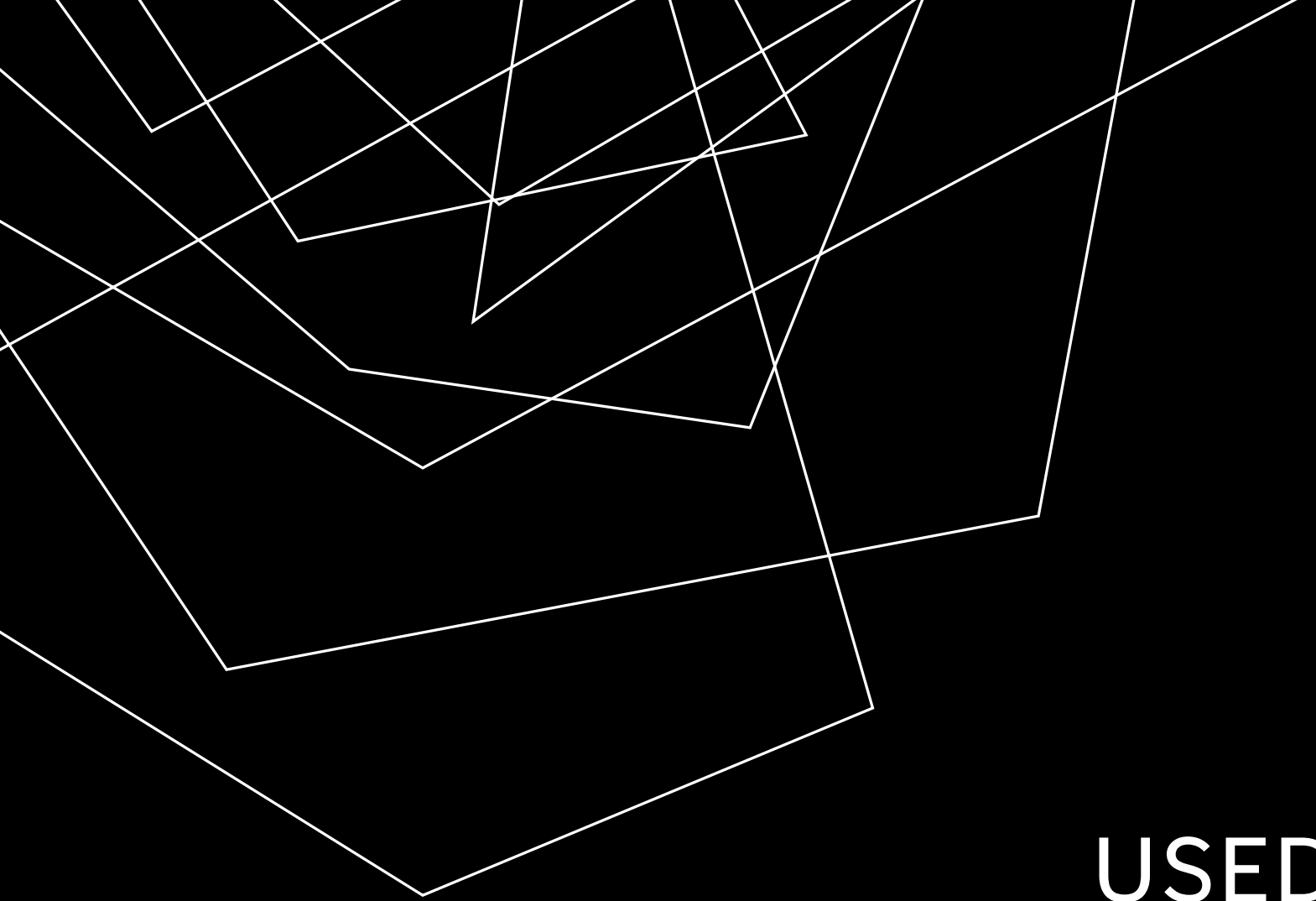


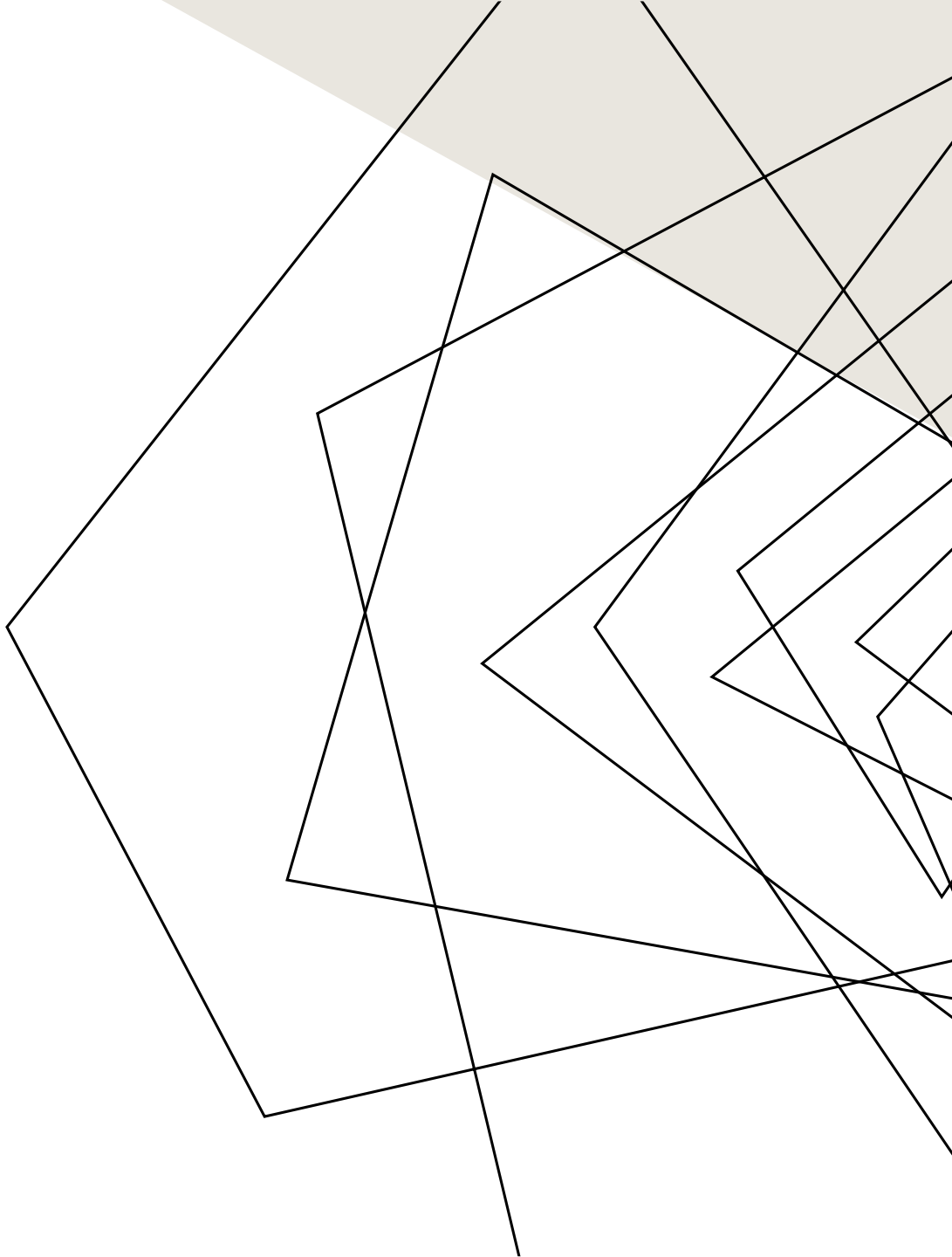
USED CAR PRICE PREDICTION

Ruoxi Hou

PURPOSE

To generate insights for both Buyers and Sellers in the European Used Car trading market;

To facilitate the purchasing and selling decision making process



PROJECT OVERVIEW

PROCESS

Data Collection

- Selenium (10250 records: Prices Web Scraping)
- PRAW: The Python Reddit API Wrapper (125562 records: Comments in subreddits)
- Manually collect images (61 brand logos)

Data Wrangling

- Currency conversion
- Missing value handling

NLP

- Sentiment Intensity Analysis

Data Storage

- .csv files
- MySQL Workbench

EDA & Dashboard

- Matplotlib
- Seaborn
- Power BI

ML

- Auto ML (Pycaret)
- XGBoost Regressor
- One-hot Encoding
- Target Encoding
- Cross Validation
- Hyperparameter tuning(Random Search& Grid Search)
- Pipeline

Web Development

- Flask

https://www.ooyyo.com/

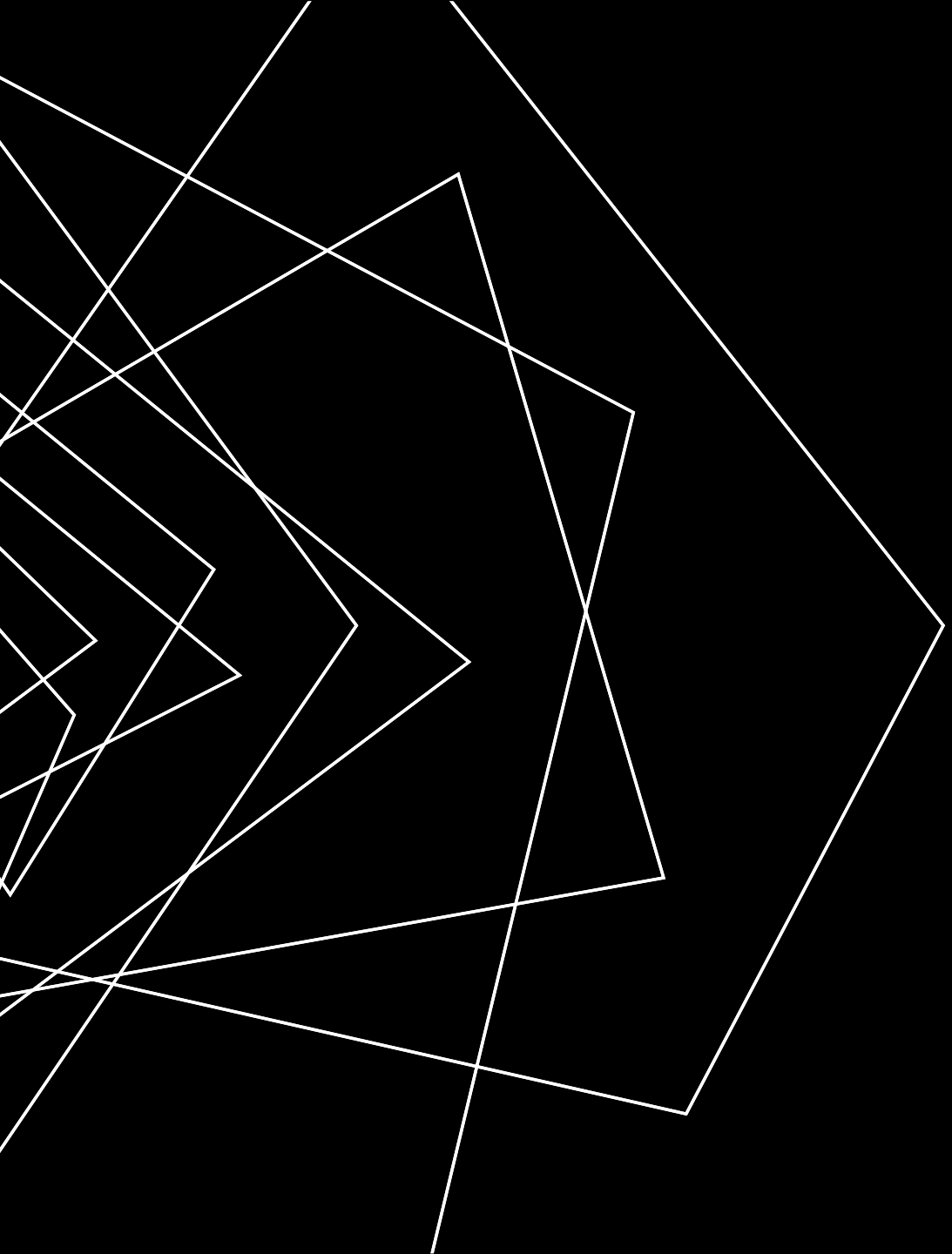
MODEL PERFORMANCE

XGBoost Regressor	**R2**	**MSE**	**MAE**
'model__bootstrap': True, 'model__max_depth': 4, 'model__max_features': 'sqrt', 'model__min_samples_leaf': 3, 'model__min_samples_split': 1, 'model__n_estimators': 1700	0.878(Training) Avg. 5 folds 0.915(Testing)	44848686(Testing)	2185(Testing)

PRODUCT PRESENTATION

FUTURE IMPROVEMENT

- **More Countries& More Features**
- **Consistent data collection for a longer time period(e.g. 10 years...) for Time Series**
- **Access to the Recommended Market Prices over the decades for Devaluation Prediction**



THANK YOU