Project Overview

Austrian Startup Analysis

Internal Lumos project to build a complete view of the **Austrian startup landscape and predict funding success**. I unified six public sources into a 500+–company dataset, enriched records with social/news signals verified using LLMs, analyzed regional/sector patterns, and trained baseline classifiers for a three class funding status.

Objective: Create a dataset of Austrian startups, model a success variable (funding vs. no-funding) and analyze regional differences in the landscape.

Approach: Ingest startup listings → extract relevant information → feature engineering (→ descriptive analytics → classification of funding status based on variables.

Scope
500+ startups
21 features
100+ working hours

Stack
Python
OpenAI API
SerperDev API

Datasets
EU-Startup Listings
Financial data
News headlines

Modeling Pipeline

Features → Model → Predict Label

Features as predictors:
- Age
- News headlines
- Region
- Category
- Business Model
- Social Media

Model Building:
- Preprocessing Pipeline
- Random Forest Model
- Cross Validation

Funding Prediction:
- inactive = startup not existing anymore
- no_funding = startup active, no funding
- funding = startup exists, got funding

Github repository

Results

500+ startups

Unified from 6 sources enriched with 21 features per company.

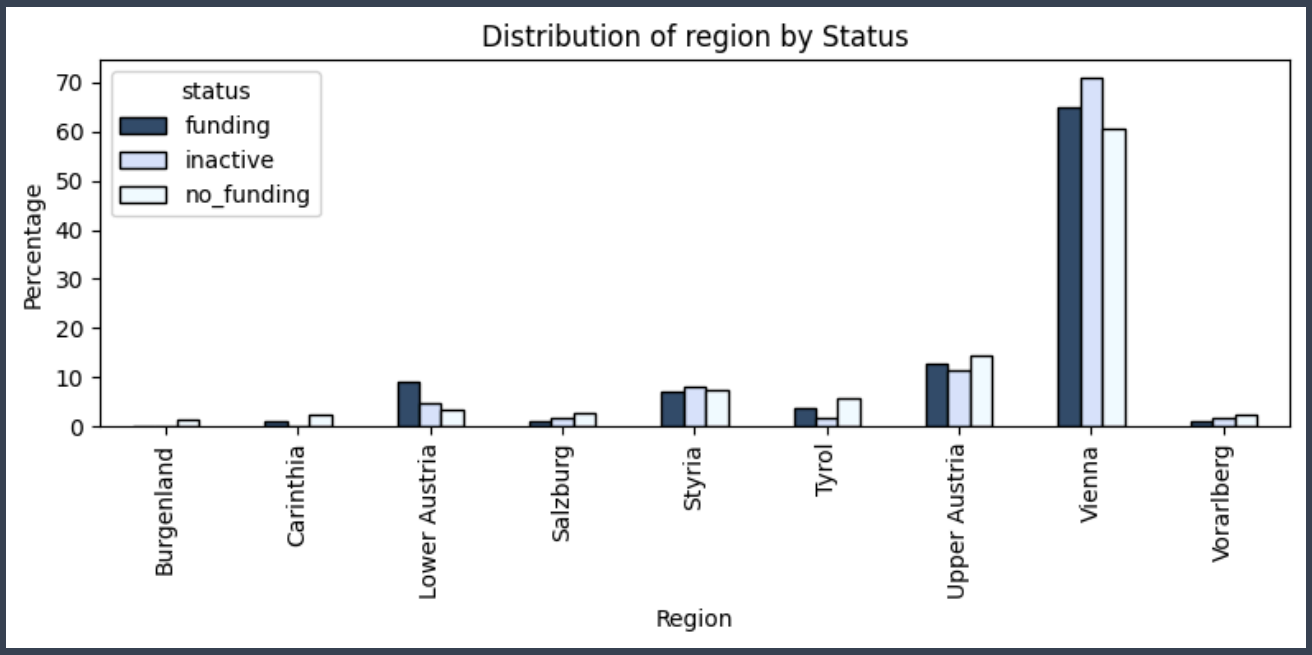
Prediction model

Built classification model to predict startup funding status.

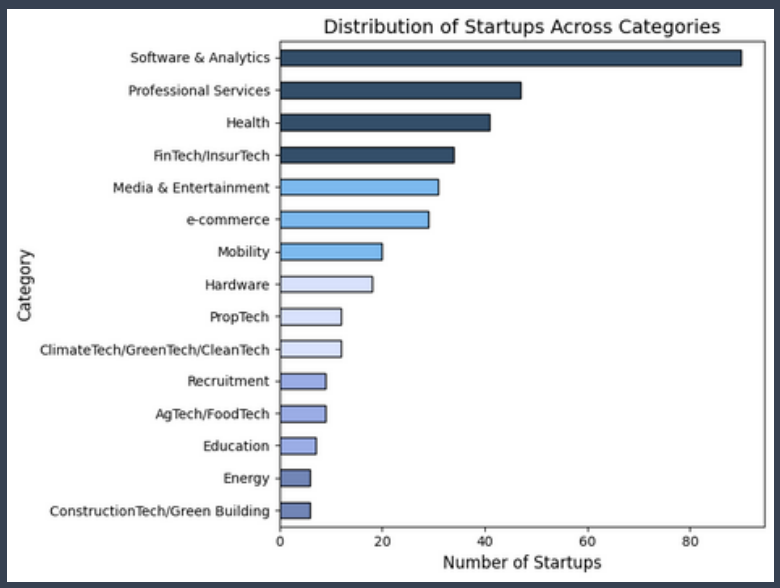
Insights

Gained insights into regional and sector specific differences.

Analysis / Modeling



Sector Distribution



Confusion Matrix