ML at Zopa

Peadar Coyle – Senior Data Scientist
Zopa’s vision

We will be the best place for money
Simple loans.
Smart investments.

Investors ➔ Invests ➔ ZOPA ➔ Repayments ➔ Borrowers

Interest + capital ➔ Loans ➔ Repayments ➔ Investors

ZOPA
<table>
<thead>
<tr>
<th>A pioneering financial services company</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>World’s 1st</strong> peer-to-peer lending platform in 2004</td>
</tr>
<tr>
<td><strong>£3 billion</strong> lent to date, and our growth is accelerating</td>
</tr>
<tr>
<td><strong>300,000+</strong> people have taken a Zopa loan</td>
</tr>
<tr>
<td><strong>60,000 +</strong> actively invest through Zopa</td>
</tr>
</tbody>
</table>
A disruptive, award-winning company
ZO PA
At Zopa we’ve invested heavily in Data

- Credit Risk assessment
- Onboarding Funnel Simulation
- Pricing Optimization
- Fraud Identification
- Document forensics
- Optimized Prospect and Existing Customer Marketing
- Income and Rent estimation for Affordability Evaluation
AI/ML Hierarchy of needs

THE DATA SCIENCE HIERARCHY OF NEEDS

LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

AI, DEEP LEARNING

A/B TESTING, EXPERIMENTATION, SIMPLE ML ALGORITHMS

ANALYTICS, METRICS, SEGMENTS, AGGREGATES, FEATURES, TRAINING DATA

CLEANING, ANOMALY DETECTION, PREP

RELIABLE DATA FLOW, INFRASTRUCTURE, PIPELINES, ETL, STRUCTURED AND UNSTRUCTURED DATA STORAGE

INSTRUMENTATION, LOGGING, SENSORS, EXTERNAL DATA, USER GENERATED CONTENT

@mrogali
Started our journey in 2014

• Hired a Chief Data Scientist

• Created a Data Science function and invested heavily in Data Analytics
How do you make ML systematic?

Wanted to be able to produce models that were:
• Rapidly generated
• Easily vettable/auditable
• Highly predictive
• Easily deployable

Several considerations
• **Common codebase** or personal choice of tools?
• Buy or **build**?
• Which language? Which package?
How do you make ML systematic? (2014)

Wanted to be able to produce models that were:

- Rapidly generated
- Easily vettable/auditable
- Highly predictive
- Easily deployable

Several considerations

- **Common codebase** or personal choice of tools?
- Buy or **build**?
- **Python**
Predictor – Zopa’s ML Toolkit (2014)

• Streamlined, Automated ML Application

• Implements all stages of producing an ML model, requiring minimal input

• Leverages PyData Ecosystem

• 9K lines
How is Predictor Used?

• **Same exact code used by DS and in Production**
  • No model translation overhead – no ‘code it all again in Java’
  • No restrictions on which ML techniques can be used

• **Training**
  • Needs unprocessed data and a simple config file
  • Driven via CLI, interacting with Python codebase, or Jupyter GUI

• **Querying Mode**
  • Driven by all of the above +
  • Rest API (Flask-based microservice) in Production, <1 -2 s/call
Learnings?

• We have full control over how we do ML
  • Methodology
  • Deployment
  • Features

• Competitive advantage, selling point to DSs

• One day we'll open source Predictor

• Need to maintain at least 2 people familiar with the code base

• Need to keep improving it – otherwise we'll miss developments
Not just ML

• We’ve been experimenting with Bayesian models, Survival Analysis, Optimization models, Natural Language Processing and Deep Learning.

• We’re strong believers that we should develop an understanding of more-than-just supervised learning

• We do regular workshops internally both to the Data team and to the wider organisation
Thank you

Further Reading

• https://blog.zopa.com/2016/10/21/the-birth-of-predictor/
• https://blog.zopa.com/2016/12/02/data-democratization/
• https://blog.zopa.com/2017/12/13/open-source-zopa/

We’re hiring Data Scientists, Risk Analysts, Product Analysts and Data Analysts
• https://jobs.lever.co/zopa/