Medi&Gene

Predictive Model

Today’s Objectives

- Dissect the data size problem
- Identify short- and long-term strategies
- Action plan

We have a data size problem

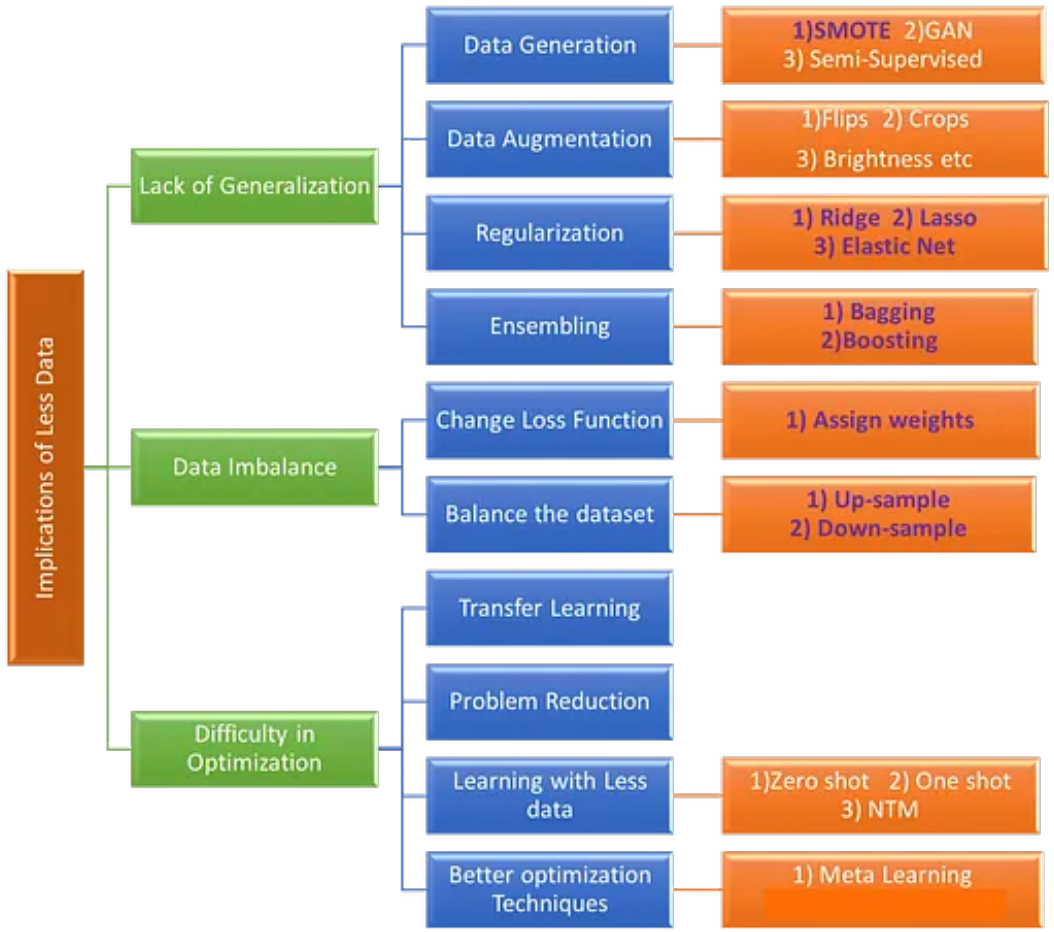
Where we are affected:

- Generalizing the solution across our sample population: 48 people may not be enough given our dataset
- Generalizing the solution across mealtimes: < 30 meals per person
- Generalizing over daily measures: ~15 days per person

Our opportunities:

- **Rich CGMS data (every 15 minutes for 15 days)**
- Statistically representative sample of the target population (early-mid 20's South Korean males & females)

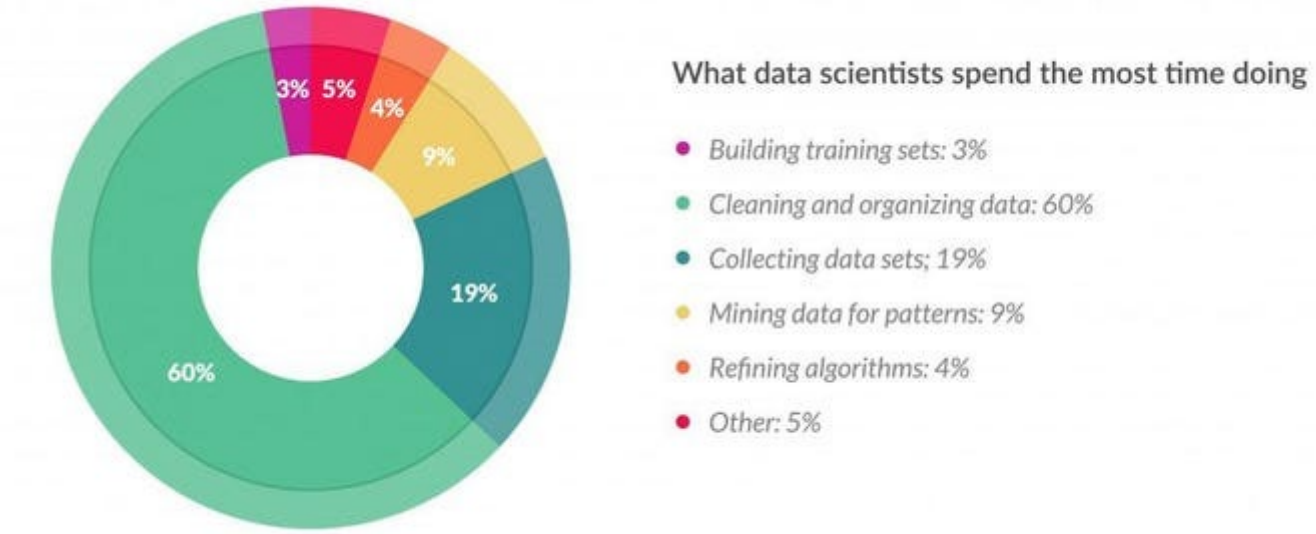
How do we handle small datasets?



Takeaways

- <u>Engineer our features/variables to optimize prediction power</u>
- Always regularize and ensemble data
- Identify opportunities for advanced techniques (transfer learning, data generation, etc.)
- Experiment with Zero-/One-shot learning for improvements

Feature Engineering – the Basics



Included in feature engineering…

Feature Selection
Missing Data Handling
Data Encoding
Normalization

Input Features for CGM Prediction Model

필수요인
- 성별
- 나이
- 교육수준
- 키
- 무게
- 허리둘레
- BMI/비만도
- 폐경
- 이상지질혈증 유병 여부
- 고혈압 유병 여부
- 흡연여부
- 식사 전 공복시간
- 식사 시간
- Total energy intake
- Total energy intake from CHO
- Total energy intake from protein
- Fiber 섭취량 / 1000 kcal
- 식품군 (15) * 수치 (5) = 75
- 음식군 (21) * 수치 (5) = 105
- 하루 평균 스트레스
- 하루 총 흡연 개피
- 하루 수면 시간
- 하루 총 운동량
- 유산소 운동량
- 무산소 운동량
- 근력 운동량
- 헬스장 운동량
- 시간별 Stress
- 흡연
- 운동

Every 15 minutes…

Demographic Context
- 성별
- 나이
- 교육수준
- 키
- 무게
- 허리둘레
- BMI/비만도
- 폐경
- 이상지질혈증 유병 여부
- 고혈압 유병 여부
- 흡연여부 (lifetime)

24-Hour Context (없는 데이터 제외)
- 하루 평균 스트레스
- 하루 총 흡연 개피
- 하루 수면 시간
- 하루 총 운동량
- 유산소 운동량
- 무산소 운동량
- 근력 운동량
- 헬스장 운동량

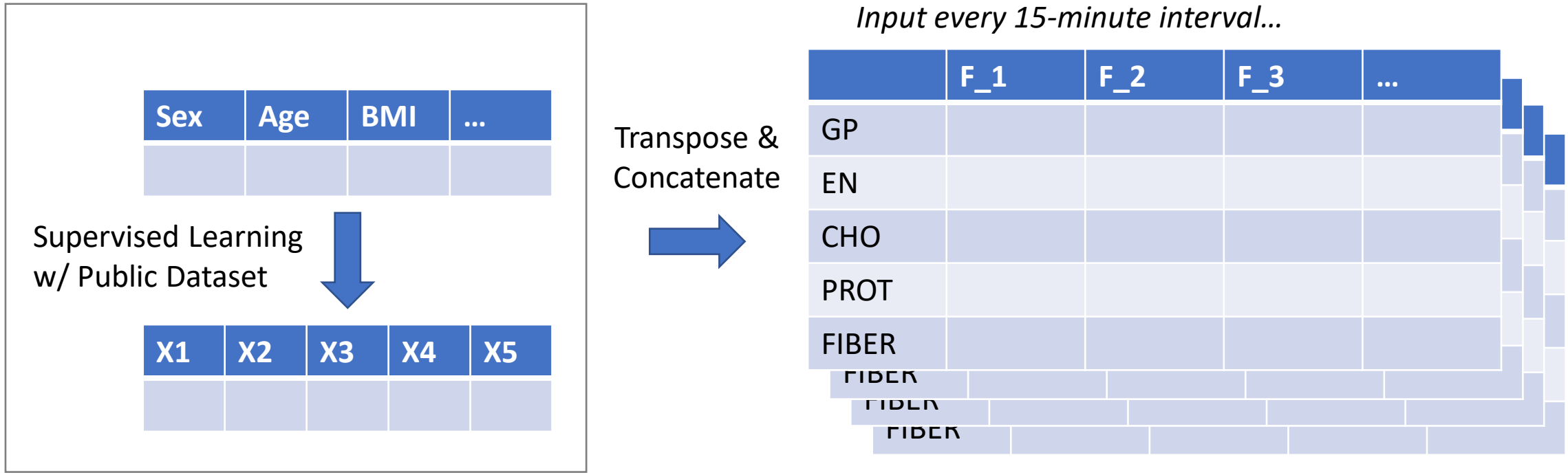
Meal Context
- 공복 시간
- 현재 시간대
- Total energy intake
- Total energy intake from CHO
- Total energy intake from protein
- Fiber 섭취량 / 1000 kcal

Food Data
- 식품군 (15) * 수치 (5) = 75
- 음식군 (21) * 수치 (5) = 105

Activity Data
- 시간별 Stress
- 흡연
- 운동

Proposal: Combining contextual data with real-time meal data

Leverage 2-D tensors as input meal data, along with contextual input provided through transfer learning



Action Plan

- Complete CNN model with Scenario 1 features
  - Expect low performance due to data size & nature of CNN with temporal data
  - Aim to hand-off prototype to an engineer (for improvements and potential re-implementation using RNN/LSTM)
- R&D + 특허내용구체화
  - GAN, Transfer Learning
  - Include zero-/one-shot learning