

# 데이터분석 세션 발표

ESC 3조



2020.8.8



TV쇼핑 [셀렉스] 매일유업 코어프로틴플  
리스 12캔+셀렉스 프로틴 키트 2박스+  
판매가 263,000원  
알뜰할인가 262,000원

무 12

매일유업코어프로틴

매일유업코어프로틴 (총 55 분)

15:35~ 16:30

# NS 홈쇼핑 데이터



TV쇼핑 [동강마루] 영월농협 햇 고춧가루  
500gX8봉  
판매가 122,000원  
알뜰할인가 121,000원

무 5

금주의 eTV 베스트

상품평 베스트



[네퓨어] 뉴네퓨어항  
사바염류 KF94마스



[삼성전자] 1등급 삼  
성 RESPOKE 냉장



[튼튼마스크] KF94  
드드마스크(대형/50



[삼성전자] 삼성 제트  
2.0 하이트 청소기/V

TV쇼핑 혜택

미리주문  
최대 10%할인  
방송 전일까지 주문시!



TV상품 쿠폰  
300원



# 1. 분석 목표

## 주제

NS Shop+편성데이터(NS홈쇼핑) 를 활용하여 방송편성표에 따른  
2020년 6월 판매실적을 예측하고,

최적 수익을 고려한 **요일별/ 시간대별 / 카테고리별 편성 최적화 방안(모형)** 제시

NS Shop+ 2019.01.01~12.31

방송일시 ▼	노출(분) ▼	마더코 ▼	상품코 ▼	상품명 ▼	상품군 ▼	판매단가 ▼	취급액 ▼
2019.1.1 6:00	20	100346	201072	테이트 남성 셀린니트3종	의류	39,900	2,099,000
2019.1.1 6:00		100346	201079	테이트 여성 셀린니트3종	의류	39,900	4,371,000
2019.1.1 6:20	20	100346	201072	테이트 남성 셀린니트3종	의류	39,900	3,262,000
2019.1.1 6:20		100346	201079	테이트 여성 셀린니트3종	의류	39,900	6,955,000
2019.1.1 6:40	20	100346	201072	테이트 남성 셀린니트3종	의류	39,900	6,672,000
2019.1.1 6:40		100346	201079	테이트 여성 셀린니트3종	의류	39,900	9,337,000
2019.1.1 7:00	20	100305	200974	오모떼 레이스 파운데이션 브라	속옷	59,000	6,819,000
2019.1.1 7:20	20	100305	200974	오모떼 레이스 파운데이션 브라	속옷	59,000	15,689,000
2019.1.1 7:40	20	100305	200974	오모떼 레이스 파운데이션 브라	속옷	59,000	25,370,000
2019.1.1 8:00	20	100808	202377	CERINI by PAT 남성 소프트 기모 릴렉스팬츠	의류	59,900	16,133,000





TV쇼핑 [셀렉스] 매일유업 코어프로틴플  
러스 12캔+셀렉스 프로틴 키트 2박스+  
판매가 263,000원  
알뜰할인가 262,000원

무 12

매일유업코어프로틴

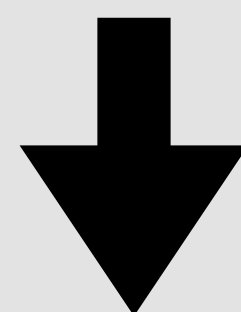
매일유업코어프로틴 (총 55 분)

15:35~ 16:30

내부요인 (고객 니즈 상품력, 시즌별 상품 변화)

+

외부요인 (날씨, 이슈, 시청률)



프로그램 매출 실적을 사전에 예측하여  
편성 최적화 모델을 찾고자 함.



TV쇼핑 [동강마루] 영월농협 햇 고춧가루  
500gX8봉  
판매가 122,000원  
알뜰할인가 121,000원

무 5

금주의 eTV 베스트

상품평 베스트



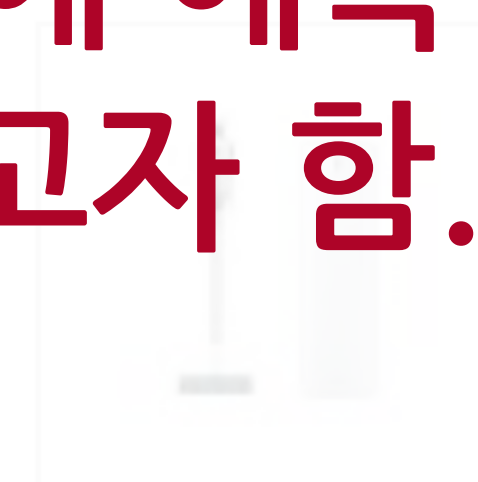
[네퓨어] 뉴네퓨어항  
사바염용 KF94마스



[삼성전자] 1등급 삼  
성 RESPOKE 내장



[튼튼마스크] KF94  
트트마스크(대형/50



[삼성전자] 삼성 제트  
2.0 하이트 청소기

TV쇼핑 혜택

미리주문  
최대 10%할인  
방송 전일까지 주문시!





## 2. 분석 데이터 특징

2019년 실적데이터 + 시청률 데이터 + 평가 데이터(2020년 6월 매출)

### [시계열 데이터]

- 방송시간 : 같은 상품군 당 20분-30분 \* 두타임 즉 1시간
- 하지만 2분, 10분, 13분 같은 적은 시간들도 존재 => **품절의 가능성이 존재하는건 아닐까?**



# 3. 분석 아이디어

## [시계열 데이터]

- 방송시간을 쪼개기. 주말(금토일)/평일로 각 24개씩 총 48개로 Embedding
- 제품별로 한 시간 단위로 묶어서 분석,  
짧은 방영시간에 관한 가중치 어떻게 따질지는 의논해봐야



## 2. 분석 데이터 특징

2019년 실적데이터 + 시청률 데이터 + 평가 데이터(2020년 6월 매출)

### [전처리 면]

- 시청률이 0임에도 불구하고 매출이 있다?
- 취급액을 판매가로 나누었을 때 정수로 나누어지지 않는다?
- 일시불/무이자 등 지불 방식이 나타나있는 항목이 있고, 아닌 항목이 있다?



# 3. 분석 아이디어

## [전처리]

- 시청률 0을 위아래 시간대 중 가장 가까운 값을 가중평균 식으로 imputate
- 판매 개수를 예측하기 위해 반올림
- 요일, 일시불, 무이자 등 다양한 변수를 학습의 새로운 변수로 활용





## 2. 분석 데이터 특징

2019년 실적데이터 + 시청률 데이터 + 평가 데이터(2020년 6월 매출)

[그 외]

- 제공된 데이터에 **코로나 시기 매출 데이터가 없어서 모델 학습 후 Accuracy 판단 어려움**
- 외부요인 중 **‘날씨’가 중요**
- 홈쇼핑의 골든타임은 **주말(금) 5시-7시**



# 3. 분석 아이디어

[그 외]

- 시청률 데이터를 온전히 고려하는 것보다는 seasonality, 요일별 분포 등 경향성 보기
- 산업군별로 증감 퍼센트가 다르니까 평균이나 분산을 추정해서 모수의 분포 적합한다음 랜덤하게 추출하여 y 값 test
- 코로나, 포스트 코로나 나누어서 모수를 예측하고, 각각 적합하기.
- 1년치 데이터를 학습하여도 우리가 정확도를 온전히 판단할 수 없으므로 분포에 가중치 주는 방법 활용



## 4. 모델 아이디어

시계열 데이터를 잘 반영할 수 있는 모델을 고민중

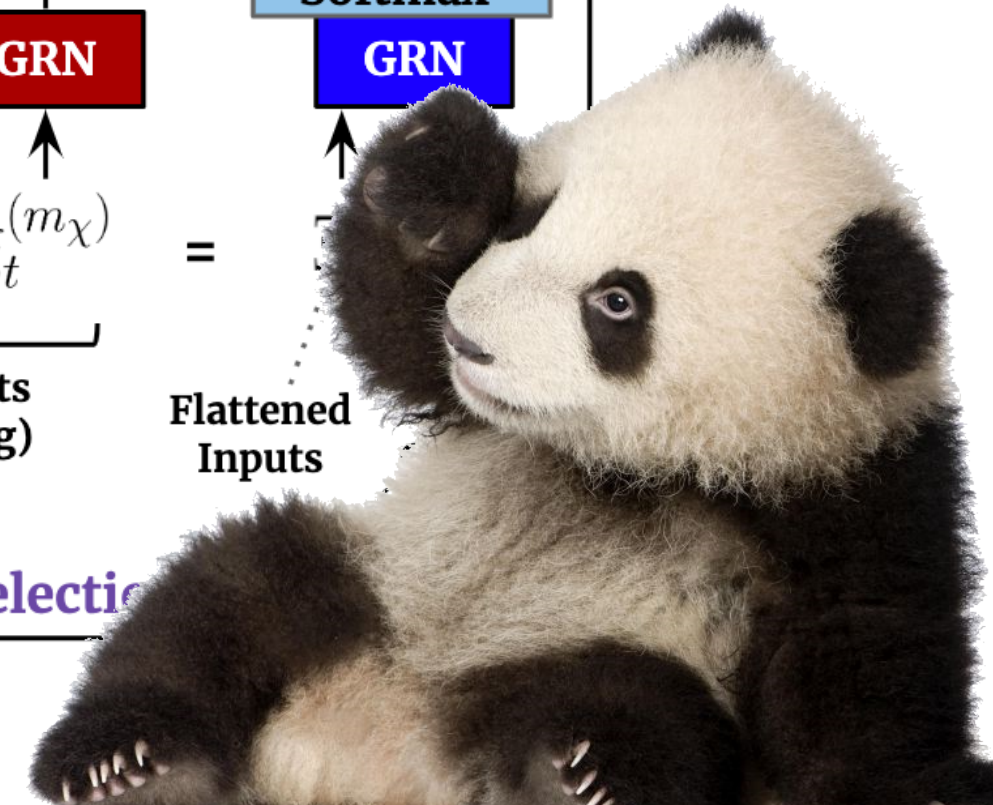
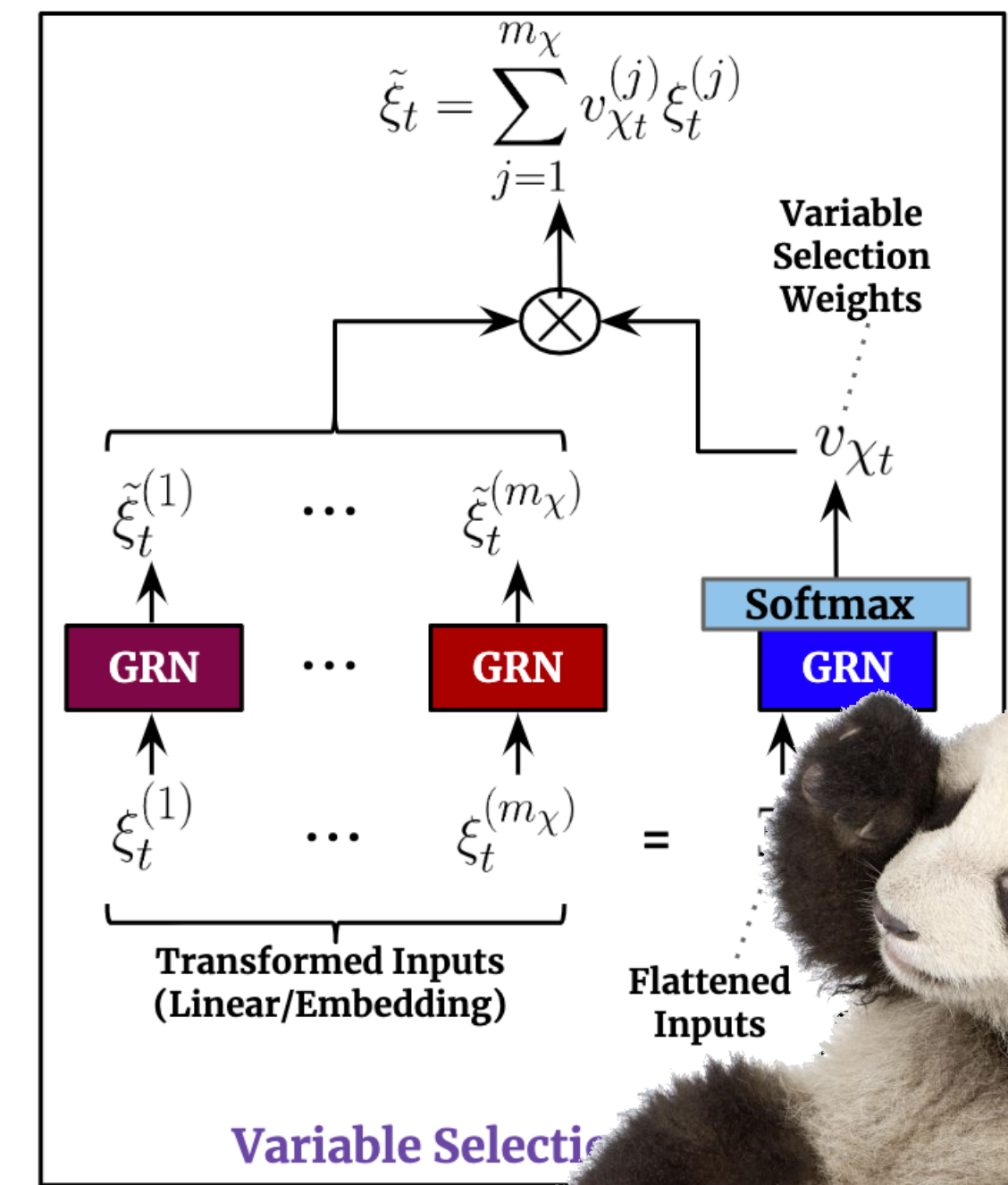
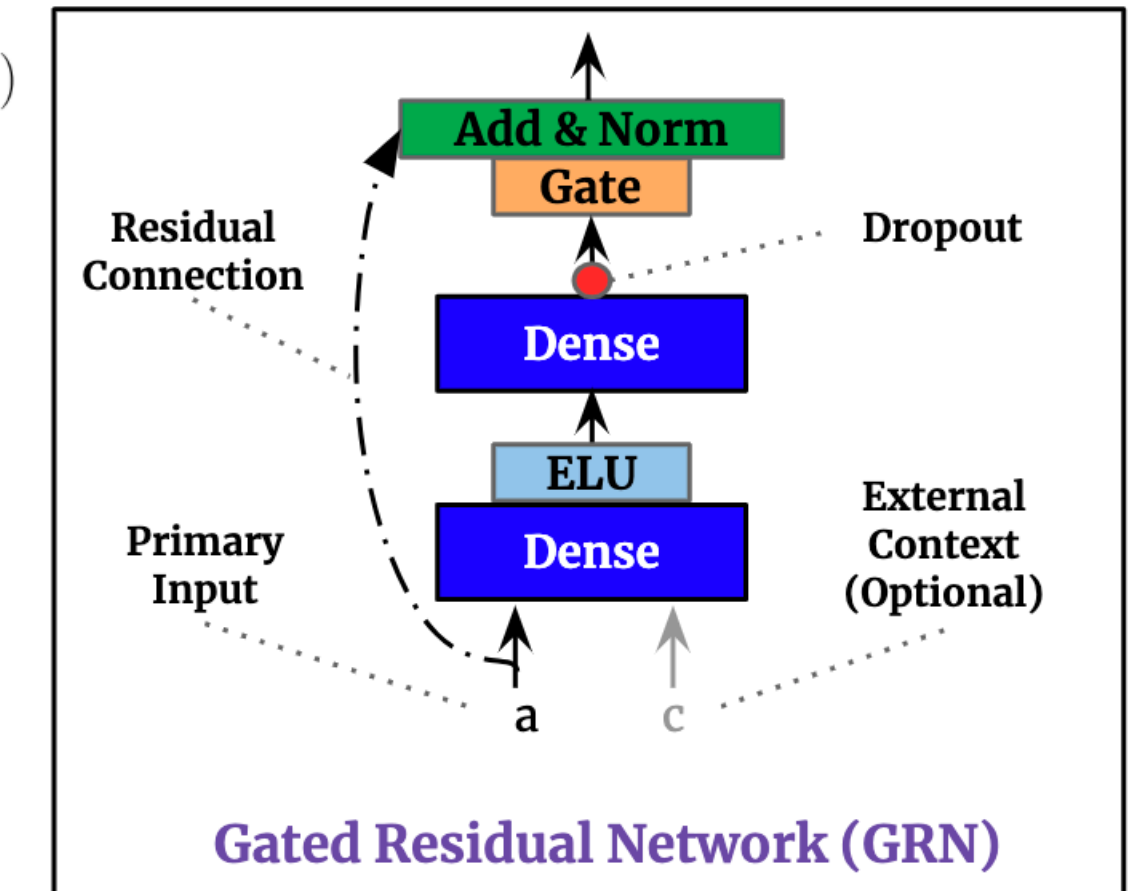
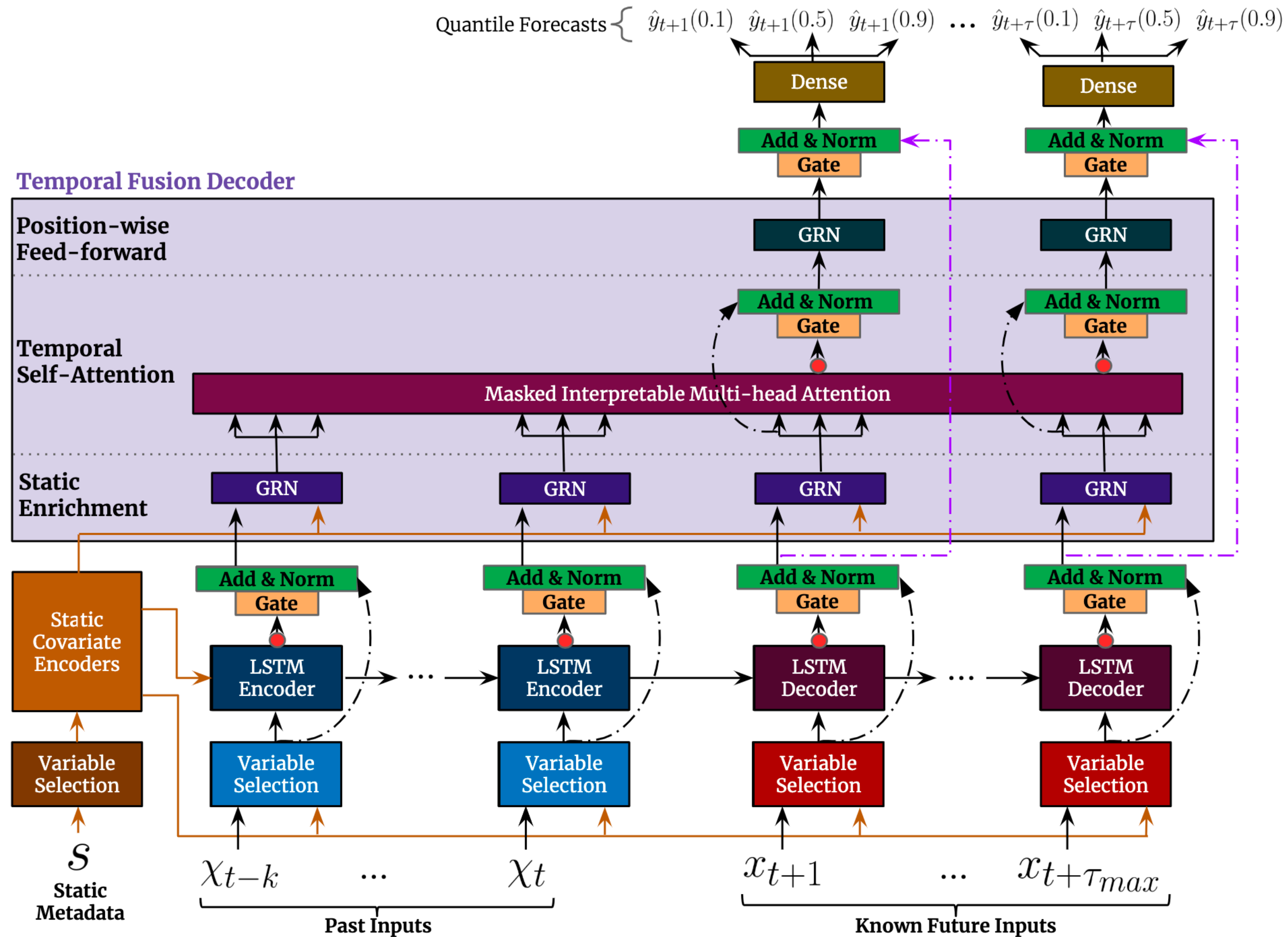
여러개의 **LSTM**을 Mixture of Experts 방법을 적용  
(각 상품 카테고리에 experts인 LSTM 사용한다는것) : 이 경우 Gating 사용

논문 읽고 어떻게 적용하면 좋을지 논의





# 4. 모델 아이디어



## 4. 모델 아이디어

**변수**, 어떤 방식으로 언제 넣으면 좋을까?

### 1. 시간 등의 변수를 embedding

Ex) 평일/주말(금) 나눠서 24개씩, 총 48개로 나누어 embedding

평일: 방영시간 (0-23)/ 주말: 방영시간 + 24(24-48) 이런 식으로

### 2. Input 변수외에도 결과값에 영향을 끼치는 외부 변수들이 있을것!

예측에 결정적인 변수들(시간, 판매량) + 메타데이터 (날씨, 계절, 상황)





감사합니다  
끝!



2020.8.8

