안녕하세요

리뷰데이터를 이용한 분석 모델을 만든 영앤이지입니다. //

먼저 팀소개입니다. /

젊고 열정가득한 저희가 데이터를 가공해 많은 사람들이 보다 쉽고/ 편리하게 생활하기를 바라는 마음에서 영앤이지라는 팀명을 정했습니다./

저희는 빅데이터/와 인공지능/을 이용해서 모든 분야에서/ 쓰일 수 있는 모델을 만들었습니다./

코알라에서 학습한/ 데이터분석 모델/ 중 솔루션 제공 모델을/ 택했고, //리뷰데이터를 이용하였습니다. /감성 분석이 필요한 제품과 서비스에 대한 고객의 만족도를 분석해 기업에 제공하려합니다/

/기존에 기업에서 리뷰를 분석할 때 많은 시간이 걸렸고, /읽는 주체에 따라 주관적으로 평가되었으며 /일일이 분류를 해야하기에 비효율적이었습니다. 이런 문제점들을/ 모두 해결할 수 있는게 저희의 모델입니다. (1분)/

아이디어 세부설명입니다. ////저희의 리뷰 분석 모델에는 두가지 모델이 들어갑니다. 저희가 데이터로 이용한 리뷰는 캐글에서 다운을 받았고, 1492개의 호텔에 대한 약 55만개의 정도의 리뷰가 있습니다.

/첫번째 모델은 리뷰에 대한 긍정과 부정을 판독하는 모델입니다. 전체리뷰를/ 모델에 넣게 되면 긍정과/ 부정으로/ 판독해 나뉘게 됩니다.

/두번째 모델은 긍정과 부정으로 분류된 리뷰의 빈도수를 통한 강점과 개선점 제시 모델입니다. 긍정과 부정으로 나뉘어진/ 리뷰들을 각각 빈도수대로 정렬을 /하여 순위를 매기게 됩니다. //이러한 분석을 제공함으로써/// 기업은 자사의 제품과 서비스에 대한 강점과 개선점을 알 수 있게 됩니다./

저희의 모델에 적용된 AI부분을 상세히 알아보겠습니다./

첫번째 모델에서는 /text classification과 random forest기법을 활용해 /긍정과 부정으로 나누게 됩니다.

코드 설명입니다.

55만개의 리뷰 중 10만개를 랜덤으로 추출합니다. /

각 리뷰에서 긍정어와 부정어의 빈도수를 세고, ratio를 측정합니다.

리뷰를 하나의 dataframe으로 합하고, 긍정은 1, 부정은 0으로 라벨링합니다.

사이킷런을 통해 train데이터와 test데이터를 분리하고, 랜덤포레스트를 이용하여 모델을 학습합니다. 그때의 정확도는 72퍼센트입니다.

정확도를 높이기 위해 데이터를 토큰화, 불용어 제거, 표제어 추출 등 전처리 후 이전의 과정을 반복합니다. 전처리를 완료한 데이터로 다시 랜덤포레스트 모델을 학습시킨 후 정확도를 비교합니다. 전처리하기 전보다 후의 정확도는 약 5퍼센트 증가하였습니다. 더 많은 양의 데이터를 전처리 후 모델 학습에 이용하면 더 높은 정확도를 얻을 수 있을 것입니다.

두번째 모델에서는 긍정과 부정으로 분류된 리뷰들을 결과에 불필요하다고 판단된 긍정어, 부정어, 불용어를 각각 없애고 자연어 처리 과정 중 part of speech tagging을 통해 각 단어의 품사를 labeling하는 작업을 합니다. /그 중 결과값에 영향을 미친다고 판단되는 /명사, 형용사, 부사의 빈도수를 추출해 기업의 //강점과 개선점을 알 수 있도록 합니다.

코드 설명입니다.

데이터의 통계 용이성을 위해 nltk를 사용하였고, 토큰화를 통한 단어의 필터링 준비 작업을 마쳤습니다. 그 후 데이터를 전처리하였고, 필요한 단어들만 필터링하여 저장하였습니다. 상위 50개의 필터링 된 단어들을 데이터프레임형식으로 저장해 시각화하였습니다.

상위에 중복된 단어들에 대해서 빈도수를 비교하기 위해, 긍정리뷰와 부정 리뷰에서 공통 단어를 추출하였고, 이를 시각화하였습니다.

이를 통해 긍정과 부정에 공통적으로 나온 단어의 빈도수를 비교하여, 이를 기업의 강점으로 판단할지 혹은 개선점으로 판단할지 알 수 있습니다. 예를 들어 room이라는 키워드는 부정의 빈도수가 긍정의 빈도수보다 높기 때문에 개선의 필요성이 강조되는 걸 알 수 있습니다.

자세한 코드는 깃허브와 flagly를 통해 올ㄹ라가있으니 꼭 확인해주시기 바랍니다.

저희 모델의 사업성 및 활용방안입니다.

4차 산업혁명시대에 알맞은 혁신적인 빅데이터 비즈니스 모델/을 통해 기업의 현 상황과 고객의 니즈를 분석해서 제시해줄 것입니다.

여러 기업의 고객 리뷰 데이터를 수집하고 분류하고 분석한 후 특정 시장에서 고객의 니즈를 파악해 전문적이고 독자적인 /시장 맞춤형 컨설팅 비즈니스를 설립하는 것이 저희의 목표입니다./

소비시장에서, 제품과 서비스에 대한 리뷰의 중요성이 나날이 확대되고 있습니다.

저희의 리뷰 분석 모델을 통해 결과를 받을 수 있다면, 기업은 리뷰의 효율적 관리가 가능해지고, 결과의 객관성까지 확보하게 됩니다. 그러므로 인적 자원의 활용성이 증가하고 리뷰를 충분히 반영했기에, 소비자 시장에서도 입지가 확대될 것이라고 예상할 수 있습니다.

홍보 방법은 크게 두가지가 있습니다.

/먼저 대형 컨설팅 기업과 협업을 하는 방식입니다. 해당업체에 컨설팅을 받는 기업에게, 저희의 모델을 이용한 감성 분석까지 포함하여, 심도있고 특색있는 컨설팅을 제공합니다.이 과정에서 컨설팅을 요청한 기업은 자연스럽게 저희의 모델을 접하게 됩니다. 이 과정에서 저연스레 우리의 모델을 홍보할 수 있습니다.

/두번째로는 기업의 전략기획과 마케팅 부서에 메일을 통해 모델을 제시하여 홍보하는 방법입니다. /감성 분석이 필요한 제품과 서비스를 주축으로 하는 기업을 타겟으로 홍보하고, 원하는 기업에 한해서 일부를 분석해 제공합니다. 분석 결과에 대해 기업이 만족을 했을 시 후에 저희의 비즈니스로 자연스레 유입될 것입니다.

초기에는 /분석 건당 수익과 협력사의 컨설팅에서 발생하는 수수료 수익으로 수익창출을 할것입니다.// 후에 구독모델로 전환하고, //기업컨설팅을 통해 정기적인 수입을 확보할 것입니다.

리뷰 분석 모델을 통해 기업은 기존 인력을 다양한 방면으로 활용이 가능할 것이며, 제품과 서비스를 전황하게 분석된 결과를 통해 발전가능성이 높아질 것입니다. 특정 시장에 대한 소비자의 니즈를 맞춤화하여 더 편리하고 만족도가 높은 제품과 서비스를 소비자가 이용할 수 있게합니다. /고로 저희의 모델은 기업과 소비자간에 우호적인고 신뢰성 있는 관계를 통해 다양한 시장 맞춤형 비즈니스 확대에 기여할 것이 분명합니다.

지금까지 영앤이지의 리뷰분석 모델을 소개해드렸습니다. 자세한 코드와 파일, 과정은 플ㄹㅐ그리와 깃허브에 있으니 꼭 확인해주세요. 감사합니다.