CONCLUSIONS PAPER

머신러닝을 활용한 금융 범죄 대응 사례

머신러닝을 구현하는 현실적이고 실현 가능한 10가지 방법

sas
THE POWER TO KNOW.

목 차

저자

Kerem Muezzinoglu SAS 사기방지 관리 고급 분석팀 수석 과학자

Carl Suplee SAS 사기방지 및 보안 인텔리전스 부문 제품 관리 이사

David Stewart SAS 뱅킹 보안 인텔리전스 솔루션 부문 이사

귀사의 사기방지 시스템은 범죄보다 앞서있습니까?

지능적인 사기 범죄자들은 금융 기관이 일반적인 사기 수법을 발견하고 차단하면 새로운 수법을 이용합니다. 더 정교하고 복잡한 계획을 세우고, 흔적을 덮기 위해 더 많은 관계자를 관여시키며, 수사망 바로 밑에서 범죄를 저지르고, 경로가 차단될 때마다 새로운 수법을 찾아냅니다.

즉 사기 수법은 끊임없이 진화하고 있습니다. 귀사의 사기방지 및 금융 범죄 시스템은 이 속도에 맞춰 진화하고 있습니까?

대부분의 기업은 여전히 규칙 기반 시스템을 기본 방어 수단으로 사용합니다. 규칙은 알려진 패턴을 밝혀내는 데에는 적합합니다. 하지만 규칙만으로는 알려지지 않은 사기 계획을 적발하거나 새로운 사기 패턴에 적응하거나 점점 더 정교해지는 금융 범죄 수법을 처리하기에는 부족합니다.

바로 머신러닝을 도입해야 하는 이유입니다. 범죄자가 테스트하고 우회하기 쉬운 규칙 기반 시스템과 달리 머신러닝은 모집단의 행동 변화에 적응하고 진화합니다.

머신러닝 시스템은 데이터에서 발견한 인사이트에 적응하는 분석 모델을 자동으로 구축합니다. 그 목표가 보다 현명한 신용 의사결정, 소매 업체의 고객 제안, 의료 진단, 사기 탐지 등 무엇이든 상관없습니다. 알고리즘은 시간이 지남에 따라 보다 정확한 결과를 제공하는 방법을 '학습'합니다. 이러한 적응형 모델링은 새롭게 떠오르는 사기 수법에 대응할 수 있는 가치 있는 자산입니다.

> 머신러닝 기술이 진보함에 따라 탐지 시스템은 금융 범죄의 새로운 패턴을 학습, 적응, 발견할 수 있게 됐습니다. 기업은 더 큰 손실이 발생하기 전에 범죄를 탐지하고 전면 방지할 수 있습니다.

새로운 머신러닝 활용법

이러한 필요에 의해 머신러닝은 이미 수십 년 전부터 활용되어 왔습니다. 새로운 점은 이제 방대한 데이터, 즉 빅데이터를 현실적인 비용으로 적용시킬 수 있다는 것입니다. 저렴한 데이터 저장소, 분산 처리, 보다 강력한 컴퓨터, 새로운 분석 툴 덕분에 진입 장벽이 낮아졌습니다.

데이터 사이언티스트들은 상용 하드웨어를 이용해 훨씬 더 방대한 데이터 세트를 훨씬 더 빠르게 처리할 수 있습니다. 이러한 분석 기법을 저렴한 컴퓨팅 플랫폼에서 엄청난 속도로 수백 또는 수천 번 반복 실행할 수 있기 때문에 모델은 이전보다 한층 더 나은 결과를 도출할 수 있습니다.

그러나 웹캐스트를 통한 약식의 설문 조사에서 응답자의 단 12%만이 자신의 기업이 현재 생산 과정에 머신러닝을 적용하고 있다고 답했습니다. 또한 약 40%는 자신의 기업이 향후 3~6개월 내에 파일럿 프로젝트를 시행하거나, 그러기 위한 계획을 세우고 있다고 말했습니다. 그러나 응답자의 절반 가량은 이 기능을 도입할 계획이 없었습니다.

아직 머신러닝을 사용하지 않는 88%가 관련 계획을 빠르게 추진하거나 재검토해야 할 이유에 대해 살펴보겠습니다. SAS 고급 분석 전문가들은 선두 금융 업체들과 협력해 다음과 같은 10가지 활용 사례를 도출했습니다.

- 케이스 조사를 위해 데이터를 수집하는 등 사람의 개입이 필요했던 **작업을 자동화**합니다.
- 규칙과 덜 정교한 분석 기술이 놓칠 수 있는 **금융 범죄 리스크를 더 많이 탐지**합니다.

체커부터 바둑까지 – 머신러닝의 진화

60여 년 전 컴퓨터 분야의 선구자인 아서 리 사무엘(Arthur Lee Samuel) 박사는 머신러닝을 '명시적인 프로그래밍 없이 컴퓨터가 학습할 수 있도록 능력을 주는 연구 분야'라고 정의했습니다. 독일 몬테소리(Montessori) 학교의 학생들이 스스로 학습하고 내적 동기 부여를 찾듯이 컴퓨터 프로그램은 데이터 전에 흔적들을 탐구합니다. 또한 명시적으로 가르침 받기 보다는, 즉 프로그래밍되기 보다는 해당 익스포져를 통해 학습합니다.

사무엘 박사는 세계 최초의 자율 학습 프로그램인 '체커 플레잉(checkers-playing)'을 고안했습니다. 이 프로그램은 40여 년 후 IBM의 '딥 블루 체스 플레잉(Deep Blue chess-playing)' 컴퓨터가 나오기 전까지 세계 챔피언을 이기지는 못했지만 훌륭한 체커 아마추어들을 뛰어넘었습니다.

SAS는 1979년 판별 분석을 위해 SAS 최초로 머신러닝 알고리즘을 도입했습니다. 1980년대 SAS와 다른 기업들은 로지스틱 회귀와 패턴 인식을 수행하는 알고리즘을 개발했습니다. 1990년대에는 의사결정 트리, 신경망, 더 정확한 예측을 위해 여러 학습 알고리즘을 결합한 앙상블 모델과 같은 접근법 등 고전적인 통계적 절차를 넘어선 발전을 이뤘습니다.

지난 10년 동안 IBM의 왓슨(Watson) 컴퓨터는 자연어 처리(NLP; Natural Language Processing)와 머신러닝을 사용해 미국 인기 퀴즈쇼 '제퍼디!(Jeopardy!)'의 역대 최고 우승자들 중 두 명을 이겼습니다. 2016년에는 구글(Google)의 알파고(AlphaGo) 프로그램이 머신러닝을 활용해 이전까지는 컴퓨터로 해결하기에 너무 복잡하다고 여겨졌던 바둑(Go)에서 세계 챔피언을 여러 차례 이기기도 했습니다.

아마존(Amazon)과 넷플릭스(Netflix) 역시 머신러닝을 이용해 사용자가 좋아할만한 콘텐츠를 추천합니다. 시리(Siri), 알렉사(Alexa)와 같은 지능형 개인 비서가 일상 생활을 돕거나 조언하고, 얼굴 인식을 통해 온라인 또는 모바일 지불을 인증할 때에도 머신러닝이 활용됩니다

머신러닝이라는 개념은 수십 년 전부터 존재했습니다. 새로운 점은 이제 방대한 데이터, 즉 빅데이터를 현실적인 비용으로 적용시킬 수 있다는 것입니다.

머신러닝의 다양한 종류

탐지를 위한 올바른 머신러닝 접근법은 기계를 학습시키기 위한 입력과 달성하고자 하는 목표에 따라 달라집니다.

지도 학습(supervised learning)을 통해 컴퓨터 프로그램은 표본 입력 그리고 이와 관련된 출력을 제공받습니다. 이때 목표는 이러한 입력과 출력을 매핑하는 일반적인 규칙을 고안하는 것입니다. 사기 탐지를 위해 지도 모델을 학습시키려면 사기 및 합법적인 활동 모두와 관련된 기록을 제공해야 합니다. 그러면 모델은 새로운 데이터에 적용될 때 사기 여부를 예측할 수 있는 기능 또는 명령 세트를 정의합니다.

비지도 학습(unsupervised learning)에서는 학습 알고리즘에 레이블, 즉 통계 조건의 종속 변수가 주어지지 않습니다. 이 알고리즘은 자체적으로 데이터에서 숨겨진 패턴을 발견하는 등 입력 안의 구조를 찾습니다. 이러한 경우, 금융 범죄를 나타내는 데이터가 무엇인지 모르기 때문에 모델은 데이터의 구조를 설명하는 함수를 작성하고, 표준에 맞지 않는 모든 것을 비정상 플래그로 지정한 다음 새롭게 입수되는 데이터에 이 지식을 적용합니다.

프로세스 자동화를 위한 머신러닝

사기 리스크 및 컴플라이언스에 대한 압박이 점점 더 커지는 가운데 문제를 해결하기 위해 단순히 더 많은 인력을 추가하는 것은 현실적으로 불가능합니다. 대신 업무를 능률화할 수 있는 방법을 찾아야 합니다. 수작업 프로세스는 매우 높은 컴플라이언스 비용을 수반합니다. 따라서 수동 또는 반복 프로세스를 자동화하며, 특히 금융 범죄를 탐지하는 데 적합한 머신러닝을 도입해야 합니다.

기업은 머신러닝을 사용해 다음과 같은 6가지 방법으로 컴퓨터가 훨씬 더 빠르게 수행할 수 있는 수작업을 자동화할 수 있습니다.

경보 및 SAR 전환율 자동 개선

기업에 가장 중요한 리스크에 초점을 맞추면서 자금세탁방지(AML; Anti-Money Laundering) 또는 컴플라이언스 그룹의 오탐지(false positive)를 줄일 수 있습니다. 금융 기관은 **AML 자동 할당(AML auto-referral)** 또는 **하이버네이션 기능(hibernation function)**을 활용해 이를 달성할 수 있습니다. 이러한 기능은 기존 경보 생성 프로세스로 시작해 광범위한 관련 정보를 활용함으로써 경보를 자동으로 개선합니다.

이러한 개선을 통해 다음과 같이 경보와 관련된 고객, 계좌, 수익자에 대한 중요한 세부사항을 추가합니다.

- 사전 케이스, 의심스러운 활동 보고(SAR; Suspicious Activity Report), 현금 거래 보고(CTR; Currency Transaction Report)
- 특정 거래, 일련의 거래, 고객, 계좌의 리스크를 평가하는 기존 스코어링 프로세스
- 법 집행 기관의 요청, 소환장, 부정적인 뉴스와 같은 외부 정보

사기 탐지, 온라인 검색, 네트워크 침입 탐지, 실시간 신용 스코어링, 이메일 스팸 필터링과 같은 애플리케이션은 모두 머신러닝의 한 종류를 활용합니다. 머신러닝은 이전 익스포져에서 얻은 지식을 활용해 알고리즘을 지속적으로 개선하고 새롭고 더 많은 정보에 기반한 액션을 수행합니다.

자동 할당 프로세스는 머신러닝을 기반으로 다음을 수행합니다.

- 모든 구성 요소를 하나로 결합합니다.
- 모든 활동 및 데이터에 스코어링 모델을 적용합니다.
- 결과 스코어를 엔티티 수준 스코어로 통합합니다.
- 리스크 평가 프로세스 및 트랜잭션 모니터링을 통해 엔티티 수준 스코어를 실행합니다.
- 확실히 임계값을 벗어나 사람에게 보고해야 하는 것을 식별합니다.

Carl Suplee SAS 사기방지 및 보안 인텔리전스 부문 제품 관리 이사는 "서프레션(suppression)과 달리 하이버네이션은 모든 활동을 고려합니다. 즉 방정식에서 아무것도 제외시키지 않습니다. 이러한 총체론적 접근법은 따로 볼 때 놓칠 수 있는 리스크를 식별할 수 있습니다. 예를 들어, 거래의 특정한 패턴이 사기를 가리키지 않아도 다른 여러 속성이나 이벤트와 함께 발생하는 경우 금융 기관에 큰 리스크처럼 보일 수 있습니다."라고 설명합니다.

"조사원이나 분석가가 완벽하게 조사할 수 있도록 하이버네이션 접근법을 사용해 모든 위험한 활동을 모으고 모든 상황을 제시할 수 있습니다. 이러한 유형의 자동 분류가 SAR 전환율을 30~50% 증가시키는 것을 확인했습니다."

케이스 조사를 위한 데이터 수집 프로세스의 자동화

분석가는 평균적으로 작업 시간의 60~70%를 기업 내외의 서로 다른 시스템 및 엔티티에서 데이터를 수집하는 데 할애합니다. 이는 보수적인 추정치로 실제는 더 높을 수 있습니다. 조사 지원에 필요한 모든 것을 찾고 수집하는 데에만 많은 시간이 요구됩니다.

Suplee는 "머신러닝은 기술을 활용해 조사와 검색을 자동화할 수 있는 중요한 기회입니다. 머신러닝 시스템은 자동으로 데이터를 조사 및 검색하고, 데이터베이스에 대해 쿼리를 실행하며, 애플리케이션 프로그래밍 인터페이스(API) 또는 웹 호출을 통해 사람의 개입 없이 외부 데이터 제공 업체의 정보를 수집할 수 있습니다."라고 말합니다.

데이터는 KYC(Know Your Customer, 고객바로알기) 시스템, 계좌 개설 시스템, 지불 시스템, 계좌 이체 시스템 등 거의 모든 곳에서 수집될 수 있습니다. 이 데이터는 사전 케이스, SAR 내러티브, 수표, 재무제표 등과 같은 이미지 또는 렉시스넥시스(LexisNexis)나 구글(Google) 지도와 같은 외부 데이터일 수도 있습니다.

Suplee는 "18개월 동안의 거래 내역이 필요하지만, 트랜잭션 모니터링 또는 케이스 관리 시스템에 이 모든 것을 로드하고 싶지는 않다면 콜을 통해 간단히 데이터를 가져올 수 있습니다. 자동화를 통해 짧은 시간 안에 프로세스를 효율화할 수 있습니다."라고 설명합니다.

"이러한 활용 사례는 하이버네이션과 같이 오탐을 줄이는 것뿐만 아니라 정보를 수집하는 데 투입되는 사람의 작업 시간을 줄이는 것과도 관련됩니다. 빠르게 성과를 도출할 수 있습니다. 일례로 한 고객은 케이스 의사결정 시간을 20~30% 단축할 것으로 예상됩니다."

복잡한 비즈니스 규칙 개발 과정의 자동화

사실상 모든 금융 의사결정은 규칙에 의해 결정됩니다. 그 기반에는 스코어가 있을 수 있지만 판정은 비즈니스 규칙 또는 규칙 세트에서 비롯됩니다. 규칙은 업계 전문성을 바탕으로 한 도메인 지식의 축적을 나타냅니다. 전통적으로 비즈니스 사용자는 필요와 관찰을 기반으로 이러한 규칙을 만듭니다.

그러나 앞서 언급했듯이 사기 범죄자들은 능숙하게 규칙을 테스트하고 우회합니다. 즉 효과적인 규칙은 세상에서 발생하는 일에 적응하고 이를 반영해야 합니다. 이때 머신러닝을 이용하면 대량의 데이터를 조사해 현황을 반영하는 규칙을 수립하는 데 도움이 됩니다.

일부 금융 채널은 리스크 스코어가 없는 경우에도 규칙에 의한 거버넌스에 특히 더 적합합니다. 예를 들어, 지불 관련 활동은 많은 당사자와 관련된 여러 범주형 속성으로 특징지어지는데 이때 각 범주는 서로 다른 리스크 수준을 나타냅니다.

Kerem Muezzinoglu SAS 사기방지 관리 고급 분석팀 수석 과학자는 "범주형 데이터를 처리하고 머신러닝을 통해 규칙을 구현하는 기본 툴은 의사결정 트리입니다. 의사결정 트리의 역사는 30년이 넘었지만 오늘날 많은 최첨단 방법의 기본 요소입니다."라고 설명합니다.

"여러 가지 방법으로 의사결정 트리를 만들 수 있습니다. 사람의 개입 없이 기계에 의해 생성된 트리는 매우 투명하며, 분리를 극대화하도록 고안된 구체적인 설계 원칙을 따릅니다. 의사결정 트리의 줄기(branch)를 따라가면 해당 의사결정에 포함되는 모든 데이터 요소를 볼 수 있습니다."

"기계는 데이터 안에서 사기가 많은 버킷을 가리키는 7개 또는 8개의 노드로 구성된 일반적인 줄기를 매우 쉽게 발견할 수 있지만, 사람에게는 매우 어려운 일입니다. 그러나 사람은 이러한 줄기를 쉽게 이해할 수 있습니다. 트랜잭션 데이터의 고도로 구조화된 관계 특성은 큰 도움이 됩니다. 의사결정 트리가 발견한 것을 따라가면서 배울 수 있습니다."

자연어 노트 및 내러티브 자동 생성

가장 간단한 형태로 자연어 생성(NLG)은 데이터를 평문으로 변환합니다. 여러 소스 및 위치(내부 및 외부)에서 데이터를 가져오고 해당 데이터를 읽을 수 있는 텍스트로 변환해 조사를 지원하고 금융 범죄 탐지 프로세스를 자동화합니다.

이 기술을 활용해 경보, SAR, CTR에 대한 노트와 내러티브를 자동으로 생성할 수 있습니다. 자연어 처리를 통해 분석가는 쉽게 콘텐츠를 읽고, 이해하고, 보완한 후 보고할 수 있습니다.

특히 이미지 인식과 결합된 자연어 처리는 문서 유형을 식별하고 이 분류를 기반으로 상황별 분석을 적용해 약 85%의 정확도를 제공합니다.

Suplee는 "자연어 생성은 기업이 반드시 고려해야 할 사용하기 쉬운 형태의 인공지능입니다. 효율성 측면에서 또 하나의 이득이기 때문입니다."라고 덧붙입니다.

인력 프로세스를 지원하는 가상 디지털 비서 개발

인공지능 기반의 SAS 감시 봇은 머신러닝을 활용합니다. 이 감시 봇은 클릭, 워크플로우, 데이터 이동, 데이터 포인트, 외부 데이터 입력, 수동 데이터 입력 등 데이터를 캡처하고, 모든 정보를 활용해 사람이 개입되는 프로세스의 다음 단계를 예측합니다.

Suplee는 "항상 스크린 A와 스크린 B를 순서대로 활용해 특정한 유형의 경보에 대한 부정적인 뉴스를 검색한다고 가정해보겠습니다. 이러한 반복적이고 예측 가능한 작업을 시작하는 대신 시스템은 특정 유형의 경보에 대한 일반적인 사용자 인터랙션을 학습하고 자동으로 제공할 수 있습니다."고 말합니다.

이 시스템은 사람을 통해 학습하고, 그 지식을 활용해 사람의 작업을 용이하게 하고 자동화합니다. Suplee는 "이 시스템은 근본적으로 이전에 사람이 했던 것과 정확히 같거나 비슷한 작업을 사람을 위해 대신해줄 뿐입니다."라고 덧붙입니다. 머신러닝은 더 나아가 과거에 행해진 것을 토대로 조사원 또는 분석가가 취해야 할 다음 단계, 즉 나아가야 할 경로를 권고할 수 있습니다.

자연어 생성

- 시스템 또는 외부 데이터 사용
- 데이터를 평문 텍스트로 변환
- 케이스 또는 SAR 내러티브 자동 생성

문서 유형의 자동 인식 및 검사

일반적인 글로벌 무역 부서의 분석가는 정보 패킷을 사용해 국제 신용장을 처리합니다. 이때 각각의 무역 패킷은 20가지 또는 30가지 유형의 문서를 포함합니다. 인지 컴퓨팅은 이러한 패킷 내에서 다양한 형식의 문서를 인식하는 지루한 프로세스를 자동화할 수 있습니다.

예를 들어, 분류(classification) 분석은 문서가 선하증권인지 송장인지를 구분합니다. David Stewart SAS 뱅킹 보안 인텔리전스 솔루션 부문 이사는 "송장으로 분류된 후에는 오버 인보이싱, 언더 인보이싱, 민군 겸용 물품 관련 조건, 기타 리스크 지표 등을 찾을 수 있습니다. 몇몇 상위 10대 글로벌 은행이 참여한 파일럿에서 약 85%의 정확도로 이 프로세스를 자동화하고 수작업을 없앴습니다."라고 설명합니다.

"또 다른 파일럿에서는 팔레스타인 보이콧 언어와 같은 것을 찾기 위해 약 9,000건의 SWIFT 메시지를 검사했습니다. 사람이 이 메시지를 검토하는 데에는 메시지당 약 5–7분이 걸립니다. 이 파일럿에서 우리는 한 메시지당 이미지 인식 및 컨텍스트 분석을 1초 이내에 완료했습니다."

이 파일럿에서 9,000건의 메시지 중 약 2.5%에 대해 몇 가지 유형의 추가 조치를 취했으며, 그 중 2 건은 제재 법률에 따라 실제 보고로 이어졌습니다. 결론적으로 이전까지 2명의 사람이 일주일간 투입돼야 했던 중요한 컴플라이언스 작업을 1분 이내에 마쳤습니다.

Stewart는 "이 기업의 규모와 인력 수준을 고려하면 자동화를 통해 천만 달러 수준의 성과를 달성한 것과 같습니다. 세계 무역 금융 문제의 측면에서 이러한 가능성은 확실히 매력적입니다."라고 덧붙입니다.

사기 및 금융 범죄 탐지를 위한 머신러닝

희귀 이벤트 탐지

희귀 이벤트를 실시간으로 탐지하려면 거래와 관련된 여러 엔티티의 모든 기록을 추적해야 합니다. 고품질의 스코어는 거래와 관련된 개인/엔티티의 개별적인 이력은 물론 관계까지 모두 관찰하고 추적해야만 생성할 수 있습니다.

Muezzinoglu는 "항상 새로운 엔티티와 관계가 발생하고 본질적으로 변화하기 때문에 모델 개발 과정에서 모든 관계와 가능한 행동을 예견하고 하드코딩할 수는 없습니다. 이력 데이터는 중요하지만, 모델 그리고 관련 탐지 시스템이 프로파일, 시계열, 이러한 엔티티의 전체 이력 및 관계를 평가할 수 있도록 가르쳐야 합니다."라고 말합니다.

"실시간 행동을 추적하는 우리의 접근법은 매우 정교합니다. 금액, 국가 코드, 기타 거래의 여러 속성을 포함해 모든 활동을 원시 시계열로 저장합니다. 모든 엔티티에 대해 이러한 속성을 추적하고 보존하며 시계열로 저장하고, 이를 기반으로 밀리 초 단위로 실시간 의사결정을 내리며 즉시 예측으로 변환합니다."

시계열 상에 방대한 처리량이 존재해도 이력을 요약하지 않고 모든 세부사항을 풍부하게 보존하는 것이 중요합니다. 동시에 사기 탐지를 개선하기 위해 다수의 요약이 계산되고 현재 이벤트와 비교됩니다.

더 많은 디지털 결제 사기 탐지

SAS는 디지털 결제 모델을 적용해 실시간 사기 탐지를 더 빠르게 구현했습니다. 이 모델은 오탐지가 거의 없이 사기의 50%를 포트폴리오의 단 0.5%에 해당하는 비용만으로 경보하는 데 성공했습니다.

이 모델은 각 송금 계좌, 수익자, 온라인 사용자를 추적해 다음과 같은 주요 지표를 확인합니다.

- **송금 및 수익자 계좌** – 수령 국가, 금액, 월 결제 당일, 송금인 및 수익자의 관계 기간
- **온라인 사용자** – 전자개인정보(Device fingerprint), 브라우저 데이터, 결제 템플릿 내역, 인증 신호

Muezzinoglu는 "이 시스템은 모든 정보는 물론 텍스트 입력을 시계열로 저장하고 이를 실시간으로 사용합니다. 그런 다음 이 시계열에서 수만 개의 변수를 생성하고 서로 다른 요약을 제공합니다. 또한 모델이 위험하다고 판단하는 특정 국가 또는 정적인 행동을 검사하는 하드코딩된 리스크 테이블을 사용합니다."라고 설명합니다.

"뿐만 아니라 우리는 추론 엔진을 구축하고 신경망, 랜덤 포레스트, 부스팅된 의사결정 트리 등을 비롯한 다양한 머신러닝 기법을 시도합니다. 일반적으로 전체 개발 사이클을 세 차례 반복해 최상의 성과를 내는 기법인 '챔피언'을 찾습니다. 그 결과 도출된 모델은 대개 수백 개의 후보를 이긴 챔피언입니다."

텍스트 분석 인사이트를 활용해 더 많은 사기 탐지

사기 탐지 모델을 구축할 때에는 모든 세부사항을 고려해야 합니다. 금액, 위치 정보, 거리와 같은 분명한 세부사항 외에도 특정 거래에는 설명과 주석이 수반됩니다.

이러한 텍스트는 사용자가 제공하는 것이기 때문에 사용자와 관련된 고유한 특성을 반영할 수 있습니다. 또한 이 텍스트는 계좌 번호나 거래 금액과 같은 수치 데이터의 특정 속성을 모호하게 할 수 있습니다. 예를 들어, 수익자가 사업체인지 개인인지, 특정 단어나 구절이 일관되게 사용되고 있는지, Co, Inc., LLC와 같은 일반적인 유의어가 개별 엔티티를 하나로 연결할 수 있는지 등의 의문이 발생합니다.

SAS 모델은 사업체와 개인을 약 80%의 정확도로 구별했으며, 일부 사례에서 탐지 능력을 약 10% 향상시켰습니다. Muezzinoglu는 "쉽게 말해 중요하지 않은 것처럼 보이는 세부사항까지도 모두 간과해서는 안 됩니다. 작은 노트가 매우 풍부한 데이터 소스의 역할을 수행하기도 합니다."라고 말합니다.

비지도 학습으로 몰랐던 사실 발견

한 대규모 글로벌 금융 기관은 SAS와 협력해 방대한 데이터 세트를 조사하고 몰랐던 사실을 발견했습니다. 예를 들어, 고객 기반에서 고위험 고객이 제대로 분류되지 않았다는 사실 등을 찾아낼 수 있습니다.

모델은 비지도 학습을 통해 방대한 데이터를 분석하고 비정상적인 사항들을 찾아내 기업 평판 리스크의 한 형태를 만들 수 있습니다. Stewart는 "이때 사람이 선한지 나쁜지를 반드시 알 필요는 없습니다. '엣지 케이스'를 대표하는 사람들, 즉 주변인과 비교했을 때 비정상적인 행동을 하는 사람들을 찾으면 됩니다."라고 설명합니다.

"우리는 최신 알고리즘, 즉 200개의 트리를 갖춘 랜덤 포레스트를 적용했으며, 현금 서비스 사업체처럼 행동했지만 온보딩 프로세스 과정에서 그렇다고 밝히지 않았거나 합법적인 현금 서비스 사업체로 등록되지 않은 것으로 간주된 고객을 분류했습니다. (약 10분 안에 종합한) 20억 건에 달하는 거래 모집단에서 416개의 의심스러운 현금 서비스 사업체, 이전에 알려지지 않았거나 등록되지 않은 89개의 MSB를 발견했으며, 추가 분류를 통해 수십 건의 생산적인 케이스를 도출했습니다."

모델은 비지도 학습을 통해 방대한 데이터를 분석하고 기업의 평판 리스크를 유발할 수 있는 비정상적인 사항들을 발견합니다.

성공을 위해 피해야 할 5가지

1. 두려워하지 마십시오. 머신러닝의 가장 간단한 형태는 자동화입니다.

사람들은 '인공지능'이라는 용어에 사로잡혀 생소하고 복잡하며 달성하기 어려운 것으로 여깁니다. 그러나 이미 많은 기업이 인공지능을 다양한 방식으로 활용하고 있으며, 거듭 설명했듯이 머신러닝은 오래 전부터 사용되어 왔습니다.

인공지능의 가장 간단한 형태는 프로세스를 자동화하는 데 도움이 되는 메커니즘입니다. 기업은 이러한 메커니즘을 도입하고 활용 방안을 모색해야 합니다. 웨비나를 통한 약식의 설문 조사에서 응답자의 약 40%는 사기 탐지, 약 10%는 경보 스코어링 자동화, 약 20%는 로봇 공정 자동화를 위해 인공지능을 적용할 것이라고 답했습니다. 대부분의 대기업은 이미 머신러닝을 활용해 이중 최소 하나 또는 두 가지를 수행하고 있습니다.

2. 인공지능을 맹신하지 마십시오.

머신러닝은 사람의 지혜와 경험을 보조하는 기술로 여겨져야 합니다. 분석 알고리즘이 발전했지만, 기계의 의사결정은 특히 금융 서비스 분야에서 완전히 자율적이어서는 안 됩니다. 예를 들어, 모델 업데이트 시기의 결정, 모델 재학습에 필요한 도메인 지식을 수집 및 통합하는 방법, 예상치 못한 입력을 평가하는 방법 등과 같은 적절한 감독이 필요합니다. 즉 사람의 지능은 이 과정에서 여전히 필수적이나, 점점 더 고차원 작업에 관여하게 됩니다.

3. 단순히 비용 절감만을 위해 머신러닝을 활용하지 마십시오.

머신러닝은 이전까지 사람의 수작업이 필요했던 프로세스와 의사결정을 자동화해 효율성을 확연히 개선합니다. 그러나 비즈니스 목표를 비용 절감 자체에만 두지 마십시오. 데이터 사이언티스트, 분석가, 조사관이 자동화할 수 있는 일상적인 업무로부터 자유로워지면 새로운 리스크를 발견하거나 알려진 리스크를 보다 철저히 연구하는 등 더 가치 있는 일에 시간을 할애할 수 있습니다.

4. 효율성을 위해 투명성을 희생하지 마십시오.

머신러닝은 가능한 한 효율적이어야 하지만 책임을 넘어서면 안 됩니다. 의사결정은 항상 투명해야 하며 검토자에게 설명할 수 있어야 합니다. 이는 머신러닝 시스템의 개발과 생산 수명 모두에 반영돼야 합니다. 투명성을 확보하는 데에는 비용이 들지만, 효율성이 어느 정도 감소하더라도 장기적인 책임은 필수입니다.

5. 인공지능 그 자체에만 목표를 두지 마십시오.

사기방지 및 리스크 관리 전문가들은 사석에서 "우리 기업이 인공지능을 도입했나? 다른 회사는 도입했는데 왜 우리는 안 하지? 인공지능을 도입해야 해."라고 말합니다.

Stewart는 "이러한 대화에서 비롯됐을 것 같은 문의를 많이 받고 있습니다. 그러나 인공지능 그 자체만을 위해 인공지능을 추구해서는 안 됩니다. 인공지능이 프로세스를 자동화하고, 보다 효율적으로 작업하며, 다른 전통적인 접근법이 놓칠 수 있는 행동을 탐지하는 데 도움이 된 가시적이며 검증된 사례들이 있습니다. 비즈니스 목표를 염두에 두고 시작하십시오."라고 조언합니다.

저자에 대하여

Kerem Muezzinoglu는 SAS 사기방지 관리 고급 분석팀 수석 과학자로 금융 사기에 대한 예측 분석 솔루션의 구축을 전문으로 합니다.

Carl Suplee는 SAS 사기방지 및 보안 인텔리전스 부문 제품 관리 이사로 SAS 금융 범죄 솔루션의 제품 방향, 전략 개발, 도메인 전문성을 담당합니다.

David Stewart는 SAS 뱅킹 보안 인텔리전스 솔루션 부문 이사로 전략 개발과 제품 관리를 책임지고 있으며, 전 세계적으로 SAS 사기방지 및 컴플라이언스 솔루션에 대한 마케팅 자문을 제공합니다.

더 자세히 알아보기

이 주제에 대한 더 자세한 내용은 sas.com/fraud에서,
SAS가 인공지능을 구현하는 방법은 sas.com/ai에서 확인하실 수 있습니다.

더 자세한 내용은 sas.com/korea/에서 확인하실 수 있습니다.