AUTOMATIZÁLT KERESKEDÉSI STRATÉGIA A VÁLLALATI FELVÁSÁRLÁSOK PIACÁN SUPPORT VECTOR MACHINE SEGÍTSÉGÉVEL

Kujbus Marcell[1]

Közgazdaságtudományi Kar, Budapesti Corvinus Egyetem

Abstract

There are existing anomalies in financial markets, which do not seem to disappear not even after their observation. One stylized fact in the Merger and Acquisition market is the following: if a company gets acquired, then its shares price increase sharply weeks before the point when the deal is announced up until some months after the acquisition. Hence, an efficient statistical model could generate significant excess return, that is not captured by the capital asset pricing model. Based on previous works, I gather financial variables to proxy the incentives, that lead to an acquisition. The Bloomberg provides data for my study. However, the data has to be cleared, so I apply the most modern tool for dealing with missing values, and to filter the outliers. 1-Norm soft margin Support Vector Machine is used afterwards to model this phenomenon. I find the optimal hyperparameters, that maximize the in-sample area under the Receiver Operating Charachteristic curve via cross-validation on a grid search. Out-sample results prove that the acquired companies are distinguished from the ones that are not along the reported variables. Kelly's formula gives us a dynamic money management tool to build a trading strategy using the outcome of the SVM. The strategy yields a significant alpha value.

1. Bevezetés

Egy vállalati felvásárlásnak (Merger&Acquisition - M&A ügyletnek) a bejelentése előtt történő előrejelzése már számos szerző érdeklődését felkeltette, ami talán annak is köszönhető, hogy egy valóban használható előrejelző modell szignifikáns prémiumot biztosítana (Palepu (1986), Crawford & Lechner (1996), és Barnes (1996) alapján). Ilyen ügylet megtörténte elméletileg akkor lehetséges, ha a két vállalat együttes értéke nagyobb, mint azok értéke külön-külön, így a felvásárló vállalat hajlandó a felvásárolandó vállalat jelenlegi piaci kapitalizációjához képest prémiumot fizetni a tranzakció létrejötte érdekében. A fizetett prémiummal arányosan így növekedni fog a felvásárlandó vállalat piaci értéke, amennyiben a felvásárlás sikeres lesz. Minél nagyobb a felvásárló vállalatok közötti verseny, várhatóan annál közelebb lesz a felvásárlandó vállalat részvényárfolyamának emelkedése a felvásárlás hatására keletkező többletértékhez.

Ha sikerülne pozitív, kockázattal korrigált többlethozammal rendelkező stratégiát konstruálni egy statisztikai modellre alapozva, az ellentmondana a hatékony piacok elmélete közepesen erős változatának Fama (1970) alapján, amennyiben a nyilvánosan elérhető pénzügyi adatokat fundamentális információnak tekintjük. Ugyanis ekkor a tisztán pénzügyi információk alapján szignálozni tudjuk a piacon megfigyelhető várható hozamokat, ami ellentmond annak a hipotézisnek, hogy az információk már beárazódtak az eszközökbe.

Stevens (1973) az elsők között alkalmazta a lineáris diszkriminancia elemzés módszerét egy véletlenszerűen kiválasztott vállalat kategorizálására aszerint, hogy vajon várhatóan felvásárlás tárgyává válik-e. A modellben szereplő változókat Stevens a következő öt csoportba sorolta: profitabilitás, likviditás, tőkeáttétel, aktivitás és egyéb változók. Az optimális modell pontossága (a helyesen kategorizált megfigyelések száma osztva az összes megfigyelések számával) 70%-os lett, aminek az 50%-os, véletlennek megfelelő arányhoz képesti eltérését Stevens (1973) t-próbával tesztelte: az eredményei szerint elvethető a pontosság 50%-kal

[1]Témavezető: Dr. Ferenci Tamás

való egyezőségét állító nullhipotézis 5‰ szignifikancia szint mellett.

Fogelberg *et al.* (1978) felhívja a figyelmet egy, a megfigyelésekkel kapcsolatos, alapvető problémára: a mintákba tartozó vállalatok, melyek a kutatás időpontjáig még nem váltak felvásárlás tárgyává, elképzelhető, hogy felvásárlás szempontjából igenis kívánatos célpontok lehetnek, ugyanakkor erről a tulajdonságról nem rendelkezünk semmiféle információval.

A szerzők új-zélandi vállalatok 86 elemű mintáját alkalmazták a korábbi cikkekhez hasonló modell illesztésére. A modell hasonló változókat tartalmaz, mint az előzőleg ismertetett munka (profitabilitás, likviditás, külső várakozások és részvényegységre jutó nyereség változékonysága), ugyanakkor pontossága alacsonyabb, mindössze 60%. A szerzők ugyanakkor hangsúlyozzák, hogy az összes változó együtthatója intuitív lett: magasabb profitabilitású, likviditású és a külső várakozások mentén magas értéket felvevő (lényegében magas P/E aránnyal rendelkező) vállalatok kisebb eséllyel válnak vállalati felvásárlás tárgyává, illetve a részvényegységre jutó nyereség magas változékonysága a felvásárlás esélyének csökkenését vonja maga után. Az első három változóval kapcsolatos eredmény szinkronban van a pénzügyi szakirodalmi részben később ismertetett, Jensen & Ruback (1983) nevéhez kötődő elmélettel, vagyis a hatékonyabb, jobb vezetéssel rendelkező cégeket kisebb eséllyel vásárolják fel, hiszen a felvásárlással együtt járó menedzsmentcserével kevesebb érték teremthető.

Rege (1984) hasonló következtetésre jut, mint Fogelberg *et al.* (1978): a diszkriminancia modell globális tesztjéhez kiszámolt F-statisztika nem jelzett szignifikanciát. Ő úgy magyarázza a pénzügyi mutatók inszignifikanciáját, hogy a vállalati felvásárlás nem azok aktuális értékétől, hanem jövőbeli várható alakulásától függ. Ugyanakkor, ahogy a szerző is kiemeli, a jövőbeli várakozásokra vonatkozó adatok nagyon nehezen érhetők el és minőségük mindig kérdéses.

Harris *et al.* (1982) az imént bemutatott négy tanulmánnyal ellentétben nem a diszkriminancia elemzés, hanem a probit regresszió módszerét alkalmazza. A probit modellnek kétféle változatát is illesztik: a fix együtthatós probit modell konstans koefficiensekkel rendelkezik, a véletlen együtthatós probit modell esetén pedig maguk az együtthatók is véletlen változók. A véletlen együtthatós probit modell pénzügyileg úgy is értelmezhető, hogy a felvásárló vállalattól függ, hogy a felvásárolt vállalat egy adott pénzügyi jellemzője milyen erős, vagy akár milyen előjelű hatást gyakorol a felvásárlás valószínűségére.

A szerzők a tévesztési mátrix kiszámítása helyett a mintán belüli felvásárolt vállalatok arányát (ami 4.8% a vizsgált 1974-1975 közötti periódusban) összehasonlítják a modell outputjával azokban a hipotetikus esetekben, ha a megfigyelés magyarázó változói éppen a felvásárolt, illetve a fel nem vásárolt vállalatok átlagos jellemzőivel egyeznek meg. Így az átlagos felvásárolt vállalat felvásárlásának valószínűsége a modell szerint 6.621%, az átlagos fel nem vásárolt vállalathoz tartozó valószínűség pedig 3.265%. Ez az eredmény a modell gyenge előrejelzési képességét mutatja, ugyanis a Tjur-féle R^2 értéke $6.621\% - 3.265\% = 3.356\%$.

Palepu (1986) logisztikus regressziót alkalmazott a bináris modell építésére. A szerző több, a korábban bemutatott cikkekkel is kapcsolatos hibára hívja fel a figyelmet: mivel a mintavételezés úgy történt, hogy egyenlő számban kerüljenek felvásárolt és fel nem vásárolt vállalatok a mintánkba (például Stevens (1973)), sérül a megfigyelések függetlenségének feltétele, ami miatt a statisztikai modellünk együtthatóira és magára a felvásárlás valószínűségére adott becslések asszimptotikusan torzítottak és inkonzisztensek lesznek. Ez alapján tehát elmondható, hogy a korábban bemutatott tanulmányok a felvásárlás állapotától függő, azonos elemű minták miatt szisztematikusan felülbecsülték egy vállalat felvásárlásának valószínűségét. Továbbá, ha a teszthalmazunk is egyenlő arányban tartalmaz felvásárolt és fel nem vásárolt vállalatot, akkor a tévesztési mátrix nem reprezentálja megfelelően a valóságos döntési helyzetekben elérhető tévesztési mátrixot. E mű nyomán törekedtem arra, hogy a felvásárolt és fel nem vásárolt vállalatok aránya az alapsokaságéhoz hasonló legyen, éppen ezért nem azonos arányban vettem mintát a két kategóriájú vállalatokból.

Palepu (1986) logisztikus regressziós modellje tesztelésére választott teszthalmaz így az alapsokasági arány figyelembevételével került kiválasztásra, mindössze 30 felvásárolt vállatot tartalmaz az 1 117-es számosságú tesztminta. A munka hasonló magyarázó változók bevonására tett kísérletet, mint a korábban bemutatott szakirodalmak, végül három változó bizonyult szignifikánsnak: a menedzsment hatékonyságát kifejező éves átlagos többlethozam (amit a szerző egy piaci indexmodellel operacionalizál), a vállalatméret és az erőforrás-növekedés aszimmetriát kifejező dummy változó. A teszthalmazon a modell pontossága 45.6%-os lett, és a ténylegesen felvásárolt 30 vállalaton belül a modell 24-et talált meg, vagyis itt 80%-os pontosság mutatható ki. Ennek ugyanakkor az az ára, hogy a modell a fel nem vásárolt vállalatok 55%-át tévesen felvásároltnak klasszifikálja.

A témában alapműnek számító munkák megszületése óta sokat változott a technológia. Ma már képesek vagyunk sokkal számításigényesebb komputációk elvégzésére. Éppen ezért a gépi tanulási eszközök jó választásnak tűnnek, amelyekkel a klasszifikációt elvégezhetjük. Kutatásomban az ezen a területen ígéretesnek tűnő Support Vector Machine-t (SVM) alkalmazom. Az SVM módszer kevésbé elterjedt, így annak elméleti hátterét részletesen is bemutatom. A modellbe bevont változókat a vállalati felvásárlások szakirodalma alapján válogattam össze, és a Bloomberg MA valamint EQS függvényei segítségével töltöttem le 2018. március 12-én. A megfigyelések száma így (23 450 fel nem vásárolt és 2 551 felvásárolt vállalat) különösen magas a korábbi munkákhoz képest (legfeljebb 1 100-1 200 megfigyelés).

Fő kutatási kérdésem tehát az, hogy lehet-e az SVM gépi tanulási eszköz előrejelzésére hagyatkozva a vállalati felvásárlások maximum 3 éves határidős piacán a piaci kockázattal korrigált hozamnál szignifikánsan többet elérni, és ha igen, akkor mi a kereskedés algoritmusa. Ez egyben ekvivalens azzal a kérdéssel, hogy hatékony-e az egyesült államokbeli felvásárlási, M&A határidős piac. Így ha a publikációban prezentált modell szignifikáns eredményt is produkál, az eredményekre tekintsünk végig szkeptikusan, és fordítsunk különös figyelmet azokra a feltevésekre, illetve limitációkra, amik szerepet játszhatnak az eredmények alakulásában.

A kutatás hátralévő része a következőképpen strukturálódik. A második fejezetben összegyűjtöm azokat a pénzügyi változókat, amelyek kívánatossá tehetnek egy céget a felvásárlására. A harmadik és negyedik fejezet bemutatja a Support Vector Machine metodológiáját. Ezután összefoglalom, hogy milyen adat állt rendelkezésemre, és azt hogy kezelem. A hatodik, hetedik fejezet kitér arra, hogy hogyan optimalizáltam a klasszifikáló algoritmust, illetve, hogy mlyen hatékonyan lehet előrejelezni az M&A ügylet létrejöttét. A nyolcadik fejezetben kifejtem a modell előrejelzésére építhető kereskedési stratégiát, míg az utolsó fejezet összefoglal. A dolgozathoz tartozó online függelék elérhető itt(ez egy link). A későbbiekben hivatkozott ábrák nevére kattintva betölt a weboldal.

2. Az akvizíció létrejöttét magyarázó változók a pénzügyi és a közgazdasági szakirodalomban

Jensen & Ruback (1983) cikke szerint a vállalati felvásárlásokat az egyes vállalatok menedzsmentjének hatékonysága határozza meg: az erőforrásokkal pazarlóan bánó menedzsmenttel rendelkező vállalatok felvásárlásával, így a hatékonytalan menedzsment leváltásával a felvásárló vállalat hatékony menedzsmentje jelentős többletértéket tud teremteni. Jensen & Ruback (1983) cikkének nyomán a menedzsment hatékonyságát az eszközarányos árbevétel, az adózott eredmény/árbevétel és EBIT/árbevétel, illetve a részvényegységre jutó cash flow segítségével operacionalizáltam.

Lewellen (1971) szerint a vállalati felvásárlások tradicionálisan elfogadott ösztönzői mellett (szinergiák a két vállalat együttes működéséből) egy felvásárlásnak akár tisztán pénzügyi ösztönzői is lehetnek: amennyiben a felvásárlandó vállalat nem tudja elérni az adósságának megfelelő optimális tőkeszerkezet szintjét, akkor a felvásárló vállalat értéket teremthet annak akvirálásával, és a tőkeszerkezet optimalizációjával. Lewellen (1971) munkája nyomán vettem be a modellbe az összes adósság/saját tőke változót.

Palepu (1986) szerint az úgynevezett növekedés-erőforrás aszimmetria is indokolhatja egy vállalati felvásárlás létrejöttét. A hipotézis szerint az alacsony növekedési potenciállal, de bőséges erőforrásokkal, vagy a jelentős növekedési potenciállal, de szűkös erőforrásokkal rendelkező vállalatok kerülnek felvásárlásra. A potenciálisan akvirálandó vállalat tehát nem rendelkezik a két tényező egyikével, amit a felvásárló vállalat segítségével tud realizálni, ami az értékteremtés forrása. A rendelkezésre álló adatok miatt csak a bőséges erőforrás indikátorát sikerült operacionalizálnom a készpénz/összes eszköz változója személyében.

Peat & Stevenson (2008) egy további lehetséges magyarázó változót vázol fel: a magas könyv szerinti érték/piaci érték aránnyal rendelkező vállalatok nagyobb valószínűséggel kerülnek felvásárlásra. Az előbbi összefüggés arra a tényre épül, hogy a magas könyv szerinti értékkel rendelkező vállalatok feltehetően nagyobb mennyiségű pénzzé tehető eszközzel rendelkeznek, ami korlátozza a felvásárló vállalat számára a potenciális maximális veszteséget.

3. A Support Vector Machine - lineárisan szeparálható mintán

A Support Vector Machine elméletét Vapnik (2013) dolgozta ki, a levezetéseket a cikke alapján mutatom be.

1. Definíció. Azt mondjuk, hogy egy *n* magyarázó változóval rendelkező minta lineárisan szeparálható, ha létezik a magyarázó változók által kifeszített térben *n-1* dimenziós maximális altér (hipersík), mely úgy particionálja a teret, hogy az egyik partícióban csak az egyik, a másik partícióban pedig csak a másik kategóriájú megfigyelésvektor tartózkodik.

A definíció közvetlen következményeként látható, hogy lineárisan szeparálható mintákat tekintve a szeparáló hipersíkok halmaza nem üres. Így legalább egy ilyen létezik, ám általában e halmaz számossága kontinuum. Egy dolgot azonban fontos kiemelni: az ilyen hipersíkok meredekségei[2] nem mehetnek egy adott véges (mintától függő) érték fölé, illetve egy adott véges érték alá[3]. Így a meredekségek *korlátos és zárt intervallumot alkotnak a valós számok terében*.

2. Definíció. Azt mondjuk, hogy egy szeparáló hipersík optimális, ha a megfigyelésvektorok hipersíktól vett minimális távolsága is maximális. Ezen távolság kétszeresét a szakirodalom *marginnak* hívja. Optimális egy hipersík, ha a hozzá tartozó margin maximális.

A lehetséges hipersíkok meredekségei a fentiek alapján kompakt halmazt alkotnak (a Heine-Borel tétel értelmében a valós tér bármelyik hatványában a kompaktság ekvivalens a korlátosság és a zártság egyszerre teljesülésével). Vegyük észre: egy hipersík adott, ha a meredeksége és egyik tengelymetszete adott. Intuitíve egyértelmű, hogy egy adott tengelyhez tartozó metszetek lineárisan szeparálható minta esetén ismét korlátos és zárt halmazt alkotnak. Így a meredekségek és tengelymetszetek Descartes-szorzatán értelmezett margin (távolság) függvény a folytonossága miatt Weierstraß tétele alapján *felveszi a maximumát*. Következtetésképp, egy lineárisan szeparálható SVM feladatnak mindig van optimális megoldása. A feladatunk megkeresni az optimumot.

A továbbiakban a $\cdot$ jel a vektorszorzásra utal, míg a „sima” skalárral való szorzást úgy jelölöm, hogy egyszerűen elhagyom a szorzásjelet. A vektorok alatt oszlopvektorokat értek, sorvektor esetén pedig kiírom a transzponálás jelét.

Nevezzünk mintának egy $\{x_i, d_i\}_{i=1}^{p}$ halmazt, ahol x_i egy adott $i = 1, 2, \ldots, p$-hez tartozó magyarázó változó vektor, d_i pedig az ugyanehhez tartozó bináris változó. A későbbi egyszerűség kedvéért tegyük fel, hogy $d_i \in \{1, -1\}$. Jelölje ω a keresett hipersík normálvektorát, és b a hipersík egyenletében lévő konstans tagot (tengelymetszetet). Vezessük be a $g : X \to \mathbb{R}$ függvényt, ahol $\forall x_i \in X$, $g(x_i) = \omega^T \cdot x_i + b$. A minta lineárisan szeparálható, ha

$$g(x_i) = \omega^T \cdot x_i + b \begin{cases} > 0 & ha\ d_i = 1 \\ < 0 & ha\ d_i = -1. \end{cases} \tag{0.0.1}$$

Ezt másképpen döntési szabálynak nevezzük. Tehát, ha egy eddig ismeretlen elem kategóriája a kérdés, akkor megnézve a g függvény általi képet dönthetünk: ha pozitív, akkor egyes kategóriájú, ha negatív, akkor pedig mínusz egyes.

Mivel tetszőleges két valós szám között található valós szám, ezért létezik olyan ε pozitív valós szám, hogy

$$\begin{cases} g(x_i) \geq \varepsilon > 0 & ha\ d_i = 1 \\ g(x_i) \leq -\varepsilon < 0 & ha\ d_i = -1. \end{cases}$$

Ezzel az ε számmal leosztva az egyenlőtlenséget (összehúzzuk a teret, avagy átkoordinátázzuk), illetve felhasználva, hogyha ω egy normálvektor, akkor annak tetszőleges skalárszorosa is az, azt kapjuk, hogy

$$\begin{cases} \overline{\omega^T} \cdot x_i + \bar{b} \geq 1 & ha\ d_i = 1 \\ \overline{\omega^T} \cdot x_i + \bar{b} \leq -1 & ha\ d_i = -1. \end{cases}$$

[2] Meredekség alatt a bázisvektorokkal bezárt szög tangensét értjük. Két dimenzióban elképzelve ez megegyezik a meredekségről alkotott képünkkel.

[3] Itt egy kis sarkítás történt. Van mikor az említett meredekség $+\infty$. Ilyenkor azonban létezik bázis, amire tekintve ez a meredekség egy véges szám. Áttérve erre a bázisra megoldottuk a problémát.

Összefoglalásképpen, egy lineárisan szeparáló hipersík lehetséges megoldás, amennyiben normálvektorai közül az egyikre igaz, hogy

$$d_i(\overline{\omega^T} \cdot x_i + \bar{b}) \geq 1,\ x_i \in X, \forall i. \tag{0.0.2}$$

Mostmár csak meg kell fogalmaznunk egy célfüggvényt, amit ezen korláton optimalizálunk, és készen is vagyunk.

Az 1. definició alapján a hipersík egy altér. Így a projekciós tétel alapján a magyarázó változók által kifeszített X vektortér előáll, mint a hipersík és annak az ortogonális kiegészítőjének direkt összege ($X = H \bigoplus H^{\perp}$):

$$x_i = x_{k_i} + \frac{\omega}{\|\omega\|} r_i,\ x_{k_i} \in H,\ \omega \in H^{\perp},\ r_i \in \mathbb{R},\ x_i \in X,\ \forall i. \tag{0.0.3}$$

Itt felhasználtuk azt a tényt, hogy H egy hipersík, így maximális altér, a direkt kiegészítője egydimenziós. Ez nem jelent mást, minthogy a normálvektora generálja: $H^{\perp} = lin(\omega)$. Tehát minden H-ra merőleges vektor $r\omega$, $r \in \mathbb{R}$ alakú. Használjuk fel az előző összefüggést döntési szabályunk (0.0.1) ellenőrzésekor!

$$g(x_i) = \omega^T \cdot x_i + b = \omega^T \cdot (x_{k_i} + \frac{\omega}{\|\omega\|} r_i) + b =$$

$$= \omega^T \cdot x_{k_i} + b + \omega^T \cdot \omega \frac{1}{\|\omega\|} \cdot r_i = 0 + r_i \frac{\|\omega\|^2}{\|\omega\|}$$

$$g(x_i) = \|\omega\| r_i$$

Ebből egyből kapjuk, hogy $r_i = \frac{g(x_i)}{\|\omega\|}$. Tehát nem csináltunk mást, mint a hipersík körüli pontokat felbontottuk két összetevőre, majd beláttuk, hogy a döntési szabályunkat (0.0.1) alkalmazva rá a g függvény általi kép csak a hipersíkra merőleges összetevőtől függ. Továbbá, egy pont hipersíktól való távolsága megegyezik a pont és a merőleges vetítése különbségének hosszával (ami a ω hosszának számszorosa), ami viszont pont a hozzá tartozó r valós szám abszolút értéke! A projekciós egyenletünkben (0.0.3) ezért osztottuk le a normálvektort a hosszával, mert így az egységnormálvektor számszorosait tekintettük.

A margin optimális nagyságának meghatározásakor (2. definíció) a hipersíkhoz legközelebb lévő pont távolsága a maximalizálandó. Legközelebb van egy pont, ha a hozzá tartozó $|r|$ minimális, tehát a lehetséges megoldásokról szóló (0.0.2) egyenlet alapján akkor, ha $r = \frac{g(x)}{\|\omega\|} = \frac{1}{\|\omega\|}$, vagy $r = \frac{g(x)}{\|\omega\|} = \frac{-1}{\|\omega\|}$. Ilyen pontokat tekintve a margin nem más, mint $m = \frac{1}{\|\omega\|} - \frac{-1}{\|\omega\|} = \frac{2}{\|\omega\|}$. Fontos még megjegyezni, hogy $g(x) = \pm 1,\ ha\ \omega^T \cdot x + b = \pm 1$.

3. Definíció. Azokat a pontokat (vektorokat), amelyekre $\omega^T \cdot x + b = \pm 1$, *support vektoroknak* nevezzük.

Emlékezzünk vissza: azt a szeparáló hipersíkot keressük, amelyhez tartozó margin maximális, vagyis amelyhez képest a pontok minimális távolsága is maximális. Így az optimalizálási feladatban $\frac{2}{\|\omega\|}$-t maximalizáljuk, de helyette egy ekvivalens formát tekintsünk:

$$\begin{cases} min: & \|\omega\| \\ & \\ & d_i(\omega^T \cdot x_i + b) \geq 1 \\ feltéve,\ hogy: & d_i \in \{-1, 1\} \\ & i = 1, 2, \ldots, p \end{cases} \tag{0.0.4}$$

A minimalizálandó függvény konvex, ugyanis normafüggvény. Továbbá, amennyiben az Euklideszi normát használjuk, akkor a célfüggvény nemcsak hogy konvex, de kvadratikus is. Ugyanolyan fontos, hogy a lehetséges megoldások halmaza, ami felett elvégezzük az optimalizálást, szintén konvex. Ez azért van, mert az ismeretlen változóink a ω vektor, és a b szám, de ezekben a korlát lineáris. Lineáris egyenes pontjai pedig konvex halmazt alkotnak.

Állítás. *Az SVM (minimum)feladat megoldása mindig globális minimumot határoz meg.*

Bizonyítás. Korábban már láttuk, hogy egy ilyen optimalizációnak mindig létezik optimális megoldása, tehát mindig van egy (legalább) lokális optimuma a feladatnak. Az SVM (minimum)feladata (0.0.4) a fentiek alapján konvex programozás, ugyanis konvex halmaz felett konvex függvényt minimalizálunk. Konvex minimum programozási feladatoknak azonban amennyiben létezik lokális minimuma, az globális minimum egyben, Tikhomirov (1996) alapján. □

Ezen állítás igaz volta teszi az SVM-et az egyik legerősebb machine learning applikációvá. Gyakorlatilag az összes többi gépi tanulási eszköz a lokális minimumok problémájával küzd, azokba előszeretettel konvergál (például a neurális hálók, vagy a k-legközelebbi szomszéd módszere is ilyen). Ez egy nagyon fontos és nagyon erős érv a Support Vector Machine mellett.

Többek között emiatt van az, hogy az SVM, mint modellépítési eszköz ritkán illeszt túl, ellentétben az említett módszerekkel.

A feladat megoldása [4]megadja nekünk a margint, illetve a support vektorokat. Geometriailag elképzelve[5], a margin sávjának határán lévő pontokat tekintjük support vektoroknak. Tételezzük fel, hogy illesztettünk egy modellt egy tanító halmazra, és most szeretnénk megjósolni egy új x megfigyelés kategóriáját. Ehhez a döntési szabályunk (0.0.1) értelmében ki kellene számítanunk a $\omega^T \cdot x + b$ értékét, és csak akkor értékelni x-et egyes kategóriába, ha ez az érték pozitív. Megmutatható, hogy ez az összefüggés az optimális megoldás értelmében csak az új megfigyelés és a support vektorok belső szorzatától függ.

4. A Support Vector Machine - lineárisan nem szeparálható mintán

A valóságban a legtöbb statisztikai adathalmaz nem fog lineárisan szeparálható mintát alkotni. Azonban ekkor mindig létezik olyan ϕ transzformáció, amelyet elvégezve a mintaelemeken, már egy nagyobb dimenziós, de lineárisan szeparálható mintát kapunk[6]. Az ilyen tulajdonságú ϕ függvényt bázisfüggvénynek hívjuk, és azt a teret, amiben lineárisan elválaszthatóvá válik a minta, *jellemző* térnek nevezzük. A jellemző térbe való transzformálás után az előző pontban leírt érvek alapján az optimalizálás ugyanúgy elvégezhető. A következőkben az úgynevezett kernel-trükköt fogom ismertetni, ami segítségével az optimalizálás kézben tartható.

4. Definíció. Azt mondjuk, hogy egy függvény *kernel* típusú, ha igaz rá, hogy $K : \mathbb{R}^2 \to \mathbb{R}$, $K(x,y) = <\phi(x), \phi(y)>$, ahol ϕ a bázisfüggvényt jelöli.

Ismerve a bázisfüggvényt, egyszerűen ki tudjuk számolni a kernelt, ha vesszük a belső szorzatukat a jellemző térbeli elemeknek. Ami azonban érdekessé teszi a dolgot az az, hogy sokszor sokkal egyszerűbb kiszámolni $K(x,y)$-t (tehát kevesebb az idő- és memóriaigénye), mint $\phi(x)$-et, $\phi(y)$-t, ugyanis ϕ dimenzionalitása nagyon nagy, akár végtelen is lehet! Talán ezt a legnehezebb megérteni, ezért nézzünk most egy példát, ahol négyzetes a jellemző tér. Legyen két megfigyelésvektor a térben $a = (a_1, a_2, \dots, a_n)$ *és* $b = (b_1, b_2, \dots, b_n)$. Ezekhez rendeljük hozzá azt a vektort, ami rendelkezik az összes elemmel, annak négyzetével, és az összes kétszeres szorzatával is:

$$\phi(a) = (1, \sqrt{2}a_1, .., \sqrt{2}a_n, a_1^2, .., a_n^2, \sqrt{2}a_n a_{n-1}, ..)$$
$$\phi(b) = (1, \sqrt{2}b_1, .., \sqrt{2}b_n, b_1^2, .., b_n^2, \sqrt{2}b_n b_{n-1}, ..)$$

Ez már most nagyon félelmetesen néz ki. Elvégezve a $< \phi(a), \phi(b) >$ skaláris belső szorzást azt kapjuk, hogy $< \phi(a), \phi(b) >= \phi(a)^T \cdot \phi(b) = 1 + \sum_j 2a_j b_j + \sum_j a_j^2 b_j^2 + \sum_j \sum_{k>j} 2a_j a_k b_j b_k + \dots$ Ezt egy sok magyarázó változóval rendelkező tanító halmazon nagyon nehéz kiszámolni, ugyanis a két jellemző térbeli vektor belső szorzatát $\mathcal{O}(n^2)$ hosszan tudjuk kiszámolni (n^2-tel arányos a műveletek számossága). Azonban algebrai átalakításokat végezve azt kapjuk, hogy $K(a,b) =< \phi(a), \phi(b) >= (1 + \sum_j a_j b_j)^2$. Így a polinomiális kernel függvényünk ki tudja számolni a belső szorzatot a jellemző térben mindösszesen $\mathcal{O}(n)$számolással! Ezt a szakirodalom a kernel-trükknek hívja. A kernel implicite definiálja azt a nagydimenziós teret, de úgy, hogy közben ő maga kézben tartható. A kutatásban a Gauss-kernelt használom: $K(x,y) = e^{-\frac{\|x-y\|^2}{2\sigma^2}}$, ami a legáltalánosabb a szakirodalomban. Jellemző terének dimenziója végtelen is lehet, és hiperparamétere a σ változó, amit körültekintően optimalizálni kell a modellalkotás során.

Vannak esetek, mikor a jellemző tér megtalálása még kerneleken keresztül is nagyon hosszadalmas és nehéz. Ekkor használjuk az úgynevezett „puha margin” SVM-et. Ez annyiban különbözik az eddigiektől, hogy a feladat korlátját (0.0.2) gyengítjük, bevezetve minden elemre egy úgynevezett ξ_i slack-változót. A korlát így $d_i(\overline{\omega^T \cdot x_i + b}) \geq 1 - \xi_i$, $x_i \in X$, $i = 1, 2, \dots, p$ alakú.

Ekkor triviálisan minden hipersík megoldás, csak a ξ_i-ket kell megfelelő méretűre állítani. Azonban ezeknek a slack-változóknak a nagyságát büntetjük egy úgynevezett C, azaz *cost* paraméterrel. Ekkor a feladatot a szokásos korlátok mellett a következő formában minimalizáljuk: $\|\omega\| + C \sum_{i=1}^{p} \xi_i^k$. A cost érték beállítása tetszőleges, ha $C = 0$, akkor semennyire sem

[4] Ami a duális feladat megoldásával lehetséges, a Lagrange-multiplikátor elv alapján

[5] lásd: online függelék, linszep.png

[6] lásd: online függelék, nonlinszep.png

büntetjük a hibákat, és minden hipersík megoldása a feladatnak, ha pedig $C \to \infty$, akkor tendálunk az eredeti „hard margin” SVM feladatához, abban az értelemben, hogy semmiféle félrekategorizálás nem engedhető meg. A k érték beállítása is tetszőleges, bár legtöbbször a $k = 1$ változatot használják, aminek a neve *1-Norm Soft Margin SVM*.

Ebben a munkában az 1-Norm Soft Margin SVM sugaras kernel változatát alkalmazom rácsmenti kereséssel meghatározva, valamint ötszörös keresztvalidációval és huszonötszörös bootstrappel kiválasztva a hiperparaméterek és a cost paraméter értékeit.

5. Az adatok és az elemzés menetének ismertetése

A dolgozat célja megbecsülni annak a valószínűségét, hogy egy tetszőleges amerikai egyesült államokbeli vállalat maximum 3 éven belül felvásárlás tárgya lesz. Ehhez szükségem volt egy lehetőleg megfelelően nagy elemszámú mintára, amire alkalmazhattam az SVM-et. A gazdasági válság óta eltelt 10 évet vettem a mintám időperiódusának, tehát adathalmazom 2008. január elseje óta felvásárolt egyesült államokbeli cégeket tartalmaz, a 2. fejezetben említett változókkal együtt. Ez pontosan 63 608 darab vállalatot jelent.

Kontrollcsoportot megválasztani nehezebb feladat volt. Az említett szakirodalmak közül sokban (példuál Stevens (1973)) úgy válaszották meg a -1-es kategóriájú adathalmazt, hogy adott napra tetszőleges vállalatokat figyeltek meg, amik eddig nem voltak felvásárolva. Nyilvánvalóan ez nem helyes, hiszen olyan vállalatoknak kellene alkotnia a kontrollcsoportot, amiket nemhogy eddig nem vásároltak fel, de *soha nem is fognak*. Így a kontrollcsoportot olyan módon választottam ki, hogy a korábbi minta időtartamát (2008-2018) visszatoltam az időben 3 évvel (2008-2015), majd letöltöttem minden év január elsejére 3000 különböző cégadatot, és megnéztem, hogy az előfordul-e az előző bekezdésben már említett cégekkel a következő 3 éven belül. Ha nem, akkor az a bizonyos vállalat a -1-es kategóriát kapta, ellenkező esetben törölve lett az adatbázisból. A mintavételezés során tartottam magam a Palepu (1986) által javasoltakhoz, miszerint törekedjünk a populációbeli arányban válogatni felvásárolt cégeket.

Avégett, hogy elkerüljem a hiányzó adatok rosszul történő helyettesítése által okozott torzításokat, a Rubin (1987) által kifejlesztett többszörös imputálás módszerét alkalmazom. A következőkben Danis (2012) cikke alapján magyarázom az általam is használt többszörös imputáció lényegét. Amikor szeretnénk egy hiányzó helyet értékkel pótolni, akkor 3 kritériumot fogalmazhatunk meg. Mivel a kiegészített adatokkal végzett elemzések révén megbízható és eredményes következtetéseket akarunk levonni a populációról, ezért muszáj megőriznünk a változók eloszlásait és a változók közti asszociációkat. Szeretnénk, hogy a kiegészített változók konfidencia intervalluma fedje a valós értékeket. Ha a lefedettség pontos, akkor az I. fajú hiba előfordulási valószínűsége is helyes. Ezenkívül szeretnénk, hogy a konfidencia intervallumok kellően szűkek legyenek, mert ezzel a II. fajú hibák lehetőségei csökkenek. Ez a technika egy szimuláción és Bayesi-alapokon nyugvó technika, ahol hiányos adatokkal rendelkező mintából $m > 1$ különböző verzió készül el, a hiányzó értékeket maximum likelihood elv alapján becsülve. Ezután a Rubin (1987) által ismertetett algoritmus szerint kombináljuk az eredményeket, ami így várhatóan az lesz, amit mi keresünk. A kutatás $m = 10$ különböző imputált mintát használ. A hiányzó értékek kezelése ezen módszerének alkalmazásához való hüvelykujjszabály az, hogy a mintában változónként maximum az adatok 30-40%-a hiányozzon, míg összeségében a minta 10-15%-a legyen hiányos.

Az imputálás elvégzése előtt azonban az outliereket kifilterezzük a tanító halmazból vertikális és horizontális outlier teszt alapján[7]. A vertikális outlierkiszűrési módszer a winsorizálás metódusát követi. A winsorizálás során a változókat külön-külön sorba rendezzük, majd a 95% feletti percentilishez tartozó értékeket a 95%-os percentilisnél lévő értékkel helyettesítjük, illetve az első 5%-hoz tartozó megfigyeléseket is az 5%-os megfigyeléssel helyettesítjük.

A horizontális outlier detektálás módszere a következő. A megfigyelések vektorokat alkotnak a magyarázó változók által kifeszített térben, és mint ilyeneknek, van hosszuk. Az Euklideszi-metrikát véve a megfigyeléseket hossz szerint sorba rendezzük. A legfelső 2.5%-ot, illetve a legalsó 2.5%-ot ebben az esetben eldobjuk a mintából. Fontos látni, hogy most nem tehetjük meg azt, amit előbb, mert a hosszakra tekintve nem tudjuk, hogy milyen előjelű változókból származ-

[7] Saját elnevezés.

hatott a kérdéses megfigyelés, így nem tudjuk pontosan megmondani, hogy mivel helyettesítsünk. Érdemes megjegyezni továbbá, hogy amennyiben a felvásárolt és nem felvásárolt cégek lineárisan szeparálható mintát alkotnának, úgy a horizontális szűrés elvégzése hatalmas hiba lenne, ugyanis pont a hatalmas hosszbeli különbségek differenciálják a kategóriákat. Azonban ennek az ellenkezőjéről az elemzés során megbizonyosodtam.

Az outlierek eltávolítása után a minta 22 197 nem felvásárolt, és 2 504 felvásárolt vállalatot tartalmazott[8]. Ez a megtisztított minta már alkalmas volt a többszörös imputálás elvégzésére, amiután 10 darab hiányzó értékkel nem rendelkező tanító halmazt kaptam végeredményül. A 10 imputált mátrixnak vettem a pontonkénti átlagát, amin felépíthettem a dolgozat hatáskörébe tartozó két modellt.

SVM alkalmazása A Support Vector Machine alkalmazásához mindenekelőtt sztenderdizálni kell a változóinkat, ugyanis a módszer érzékeny a mértékegységekre. Sztenderdizálás után minden szórásegységben van kifejezve, így az SVM alkalmazható. Az SVM-hez tartozó Gauss-kernel σ és a C hiperparamétereinek helytelen kiválasztása komoly problémákat okozhat, ugyanis okozója lehet a túl-, illetve alulillesztésnek is. Ezen ok miatt, két különböző módon is optimalizálom a paramétereket, majd mindkettővel elvégzem a validálást, és a jobb eredmény lesz az SVM modell végső eredménye. Mindkét módszer során a Hsu *et al.* (2003) által tanácsolt rácsmenti keresésből indulok ki, ami a lehetséges paraméter-párok egy durva felbontásán halad át, majd egyenként kiértékelve a modell illeszkedését, kiválasztja a legjobbat. A szakirodalom alapján exponenciálisan növekvő sorozatokat érdemes megadni a lehetséges értékeknek, így én a σ értéket a $2^{-6}, 2^{-5}, \ldots, 2^{2}$ számok közé szorítottam, míg a büntető paramétert a $2^{0}, \ldots, 2^{8}$ halmazból kerestem ki[9]. A modellalkotás során nem feltétlenül érdemes elérni egy nagyon magas tanuló-pontosságot, hiszen az könnyen a túlillesztés proxyja lehet. Mégis, a rácsmenti keresés során ezt maximalizáljuk, mert másképp nem tudnánk optimálisat választani a paraméterek közül. A túlillesztés detektálására azt a szabályt alkalmaztam, hogyha a tartóvektorok száma a mintaelemszám több, mint 10-15%-át teszi ki, akkor igen valószínű a túlillesztés bekövetkezte, bár ez csak egy egyénileg megfogalmazott határ.

A két különböző módszerrel optimalizált modell közül a következőképpen választottam. Az összes vállalatot tartalmazó mintát felosztottam egy, a cégek 70%-át tartalmazó úgynevezett tanuló/tanító halmazra, és egy 30%-nyi megfigyelést tartalmazó validáló mintára. A tanító mintán végeztem el a paraméterek kalibrálást, aminek a hatását a validáló mintán teszteltem. Az a modell a jobb, aminek nagyobb a validáló halmazon vett ROC-görbe alatti területe.

6. Keresztvalidált és bootstrappelt hiperparaméter optimalizálás

Az első applikáció Min & Lee (2005) alapján történik. Ennek az implementálása remekül megvan oldva az R *caret* nevű csomagjában, így ezt használtam. A lehetséges (σ, C) paraméterek mind a 64 rácspontjának az esetében külön-külön felosztottam a tanító halmazt 5 véletlenszerűen kiválasztott, különálló részre. Minden rácspont mellett elvégeztem a modell kalibrálását 4 darab ötödön, majd az ötödiken teszteltem azt. Ezután másik 4 ötödön tanítottam, hogy a maradék egyen tesztelhessek. Összeségében 64 rácspontra külön-külön 5 modellt számítottam ki, ami 320 modellt jelent. Egy (σ, C) páros eredménye így az 5 különböző modell 5 különböző tesztelő halmazán elért hatékonyságnak a számtani átlaga. Ezt elvégeze az összes lehetséges (σ, C) pároson kiválasztottam, *keresztvalidáltam* a legjobban teljesítő hiperparaméterpárt, tehát a 64 átlagolt pontossági értékek közül kiválaszottam a maximálisat.

A gyakorlatban állandóan felmerülő kérdés, hogy szeretnénk megbizonyosodni egy adott statisztikai paraméterre adott becslésünk bizonytalanságáról. Szakdolgozatomban például az SVM modell koefficiensei, illetve maga a hatékonysági mérőszámok is ilyen paraméterek. A bootstrap módszert alkalmazom az említett paraméterek bizonytalanságának számszerűsítésére (vagyis konfidencia intervallum konstruálására),

[8] A változók eloszlásait lásd az online függelék sűrűségek.png képén

[9] lásd:Min & Lee (2005)

ehhez viszont röviden ismertetem a módszer elméleti hátterét. Ehhez Efron & Tibshirani (1994) könyvének 6. fejezetét használom fel.

Kiindulópontunk egy n elemű minta, amiről feltételezzük, hogy a mintaelemek ugyanabból az eloszlásból származnak: $\mathbf{x}$=$(x_1, x_2, ..., x_n)$. A valószínűségeloszlás egy paraméterét szeretnénk megbecsülni, formálisan: $\Theta = t(F)$, ahol F az eloszlásfüggvény, t pedig az eloszlásfüggvényből számolt paraméter. Efron és Tibshirani egy egyszerű példán szemlélteti a bootstrap módszert, nevezetesen válasszuk a t paramétert a populáció várható értékének. Ekkor ismert, hogy a mintaátlaggal az eloszlás várható értékére torzítatlan becslésünk lesz, formálisan $\hat{\Theta} = s(\mathbf{x})$, ahol $\mathbf{x}$ az n elemű mintánk, s pedig a mintaátlagolás (n-változós és skalár értékű) függvénye, és $E(\hat{\theta}) = \theta$. Annak ellenére, hogy ez a becslés torzítatlan, még nem tudunk semmit a becslés bizonytalanságáról, varianciájáról. Erre a kérdésre ad választ a bootstrap módszer: válasszunk ki visszatevéses mintavételezéssel B darab n elemű mintát $\mathbf{x}$ elemeiből (mindegyik elemet egyenlő valószínűséggel választjuk), a kapott n elemű mintákat jelölje $\mathbf{x^1}, \mathbf{x^2}, ..., \mathbf{x^B}$. A kapott minták mindegyikére számítsuk ki az s becslőfüggvény numerikus értékét (a példánkban ez az átlag) , és számítsuk ki a kapott B darab átlag tapasztalati szórását. A kapott szórás megadja az eloszlás paraméterére adott becslésünk szórásának bootstrap becslését, ami a paraméter becslésére vonatkozó bizonytalanságnak egy intuitív jellemzője. A tanuló (70%-ot kitevő) halmazon a bootstrappelést $B = 25$-szor végeztem el.

A modellezés eredményét a következő táblázat foglalja össze[10], ami alapján a keresztvalidált modell a tanuló halmazon jobban teljesített, így validálásra az lesz használva.

	Keresztvalidáció	Bootstrapp
(σ, C)	$(0.5, 16)$	$(0.25, 16)$
s.v arány	20.143%	42.068%
pontosság	94.613%	94.504%
kappa	0.645	0.638
AUC	0.963	0.935

Az átlagos imputált teszt minta 7410 darab vállalatán a Support Vector Machine keresztvalidált algoritmusa **93.1%**-os AUC értékkel rendelkezik. A kereskedési stratégia megkonstruálása ez alapján ígéretesnek tűnik.

7. Kereskedési stratégia kialakítása

A bevezetőben a következő kérdést fogalmaztam meg. Lehetséges-e a modell előrejelzésére kereskedési stratégiát építeni, valamint ezzel a stratégiával a kockázattal korrigált piaci hozamnál szignifikánsan magasabb prémiumot elérni? Erre a kérdésre igenlő a válasz. A dolgozatban megadok egy hosszú távú, a vagyon növekedési rátáját maximalizáló stratégiát.

Ezen dolgozat elején említettem, hogy a felvásárlási ügylet többletet termel, így a felvásárolt vállalatok részvényárfolyama emelkedik egy ilyen bejelentés hatására. Az információk időben elnyúlva épülnek be az árakba, így a felvásárlás előtti hónapokban már elkezdenek emelkedni az árfolyamok, egészen a deal bejelentése utáni hónapokig. Barnes (1996) szerint egy felvásárolt cég átlagosan 31.25%-ot értékelődik fel a bejelentés napjához képest 2 hónappal korábbi naptól a bejelentés utáni 3. hónapig. Crawford & Lechner (1996) szerint a bejelentés előtti 50. naptól a célvállalat kivezetéséig várhatóan 39%-os emelkedés figyelhető meg. Dolgozatomban azonban ragaszkodom a Palepu (1986) által közölt számokhoz, ami 20.98% a felvásárlás, és -1.62% a felvásárlás félrekategorizálása esetén. Ezzel azt szeretném elérni, hogy egy konzervatív becslést adjak az elérhető nyereségre: ha a Palepu (1986) féle számok alapján is jövedelmező a stratégia, akkor szinte biztosak lehetünk benne, hogy a valóságban is használható az itt közölt módszer.

A Kelly-kritériumról dióhéjban A Kelly-féle befektetési stratégia arra keresi a választ, hogy ha befektetési döntések végtelen sokaságával állunk szemben, és ezeket a döntéseket időben egymás után ugyanannak az egyénnek kell meghoznia (tehát a korábbi nyereségek, veszteségek befolyásolják az aktuális vagyonszintet), akkor vagyonunk mekkora hányadát érdemes befektetni a kockázatos befektetési lehetőségbe. Természetesen a kérdés így még nem jól definiált, így Kelly Jr (2011) alapján meghatározunk egy célfüggvényt, aminek optimalizálása valóban egy olyan stratégiához vezet, ami a hosszú távú növekedési rátát maximalizálja 1 valószínűséggel. Tegyük fel, hogy diszkrét időpontokban van lehetőségünk befektetni, minden körben dönthetünk, hogy vagyonunk mekkora részét kockáztassuk. A befektetési lehetőség p való-

[10] grid search:online függelék gridsearch.png, AUC: online függelék roc_train.jpg.

színűséggel c_1 hozamot termel, és c_2 hozamot bukunk el, ha a befektetés nem jár sikerrel. A vagyonunk értéke így n kör után $W(n) = W(0) \cdot (1 + f \cdot c_1)^k \cdot (1 - f \cdot c_2)^{n-k}$, ahol f a döntésünk eredménye: az aktuális vagyonunk mekkora részét fordítsuk a kockázatos eszköz vásárlására (a maradék $1 - f$ részt készpénzbe helyezzük, tehát ez a rész nem változik), a k pedig az n kör alatt bekövetkezett nyerések száma. Egyenletünkben k binomiális eloszlású valószínűségi változó p és n paraméterekkel, így nem triviális, hogy hogyan is tudjuk ezt a kifejezést optimalizálni. Ha mindkét oldal logaritmusát vesszük, és a logaritmizált vagyont leosztjuk a periódusok számával, akkor az így kapott kifejezés végtelenben vett 1 valószínűségű határértéke a nagy számok erős törvénye szerint éppen a kifejezés várható értékéhez konvergál. Az így kapott célfüggvényt, mint f függvényét már tudjuk optimalizálni. Az optimális befektetési stratégia

$$f^* = \frac{p \cdot (\frac{c_1}{c_2} + 1) - 1}{\frac{c_1}{c_2}} \tag{0.0.5}$$

A stratégiáról belátható, hogy a $lim_{n\to\infty} \frac{1}{n} \cdot ln(W(n))$ limeszt 1 valószínűséggel maximalizálja, amit nevezzünk hosszú távú növekedési rátának Kelly nyomán. Sőt, a döntéshozó következő körhöz tartozó vagyona logaritmusának várható értékét is maximalizálja a stratégia, ez könnyen látható a hosszú távú növekedési rátával ekvivalens $p \cdot ln(1 + f \cdot c_1) + (1 - p) \cdot ln(1 - f \cdot c_2)$ képletből. Ez az eredmény úgy is értelmezhető, hogy ha a döntéshozó a logaritmikus hasznosságfüggvénnyel rendelkezne és azt körről körre mohón optimalizálná, akkor is a hosszú távú növekedési ráta maximalizálására jutna. Peters & Gell-Mann (2016) fizikusok egy mélyebb értelmezését adják az alkalmazott logaritmikus transzformációnak: ha a vagyonunk ugyanis 0-ra csökken, akkor végleg ott is marad, és a hosszú távú növekedési ráta 0 lesz, így az optimalizáláshoz egy olyan függvényre van szükségünk, ami a 0 értékű vagyont, mint elnyelődési állapotot végtelen mértékben bünteti, aminek a logaritmusfüggvény eleget is tesz. Ez röviden azt jelenti, hogy a Kelly-stratégiát használva nem mehetünk csődbe.

Thorp (2011) megmutatta, hogy amennyiben a számunkra kedvező esemény bekövetkezési valószínűsége körönként változik, akkor is a 0.0.5 képlet alapján kell kockáztatnunk a vagyonunkat a hosszú távú növekedési ráta maximalizálásához. A kutatásom keretei közt az SVM mondja meg a sikeres kimenetel valószínűségeit az egyes vállalatokra.

A Kelly-stratégia a következőképp konfigurálódik. A c_1, c_2 paraméter értékét csak *a priori* információkkal tudom meghatározni, így annak képletébe a hozamok várható értékét helyettesítem:$\frac{c_1}{c_2} = \frac{20.98\%}{1.62\%}$, Palepu (1986) cikke alapján. A p paraméter értéke vállalatfüggő, és az SVM modellnek az adott megfigyelésre vonatkozó becslésével egyenlő, vagyis, hogy mekkora eséllyel lesz felvásárlás tárgya az adott megfigyelés. A dolgozatban a shortolást nem engedem meg, így értelemszerűen csak olyan céget vizsgálok, amibe érdemes befektetni, aminek a várható hozama pozitív lesz. Ez azt jelenti, hogy a felvásárlás becsült valószínűségének ki kell elégítenie a $p \cdot 20.98\% - (1 - p) \cdot 1.62\% \geqq 0 \iff p \geqq 7.17\%$ egyenlőséget. Továbbá a Kelly-kritériumnak van egy fontos feltétele: két tét megrakása közt ki kell derüljön a fogadás sikeressége. Esetünkben 3 éves időperiódust vizsgálok, és amennyiben el akarjuk kerülni azt, hogy 3 évente mindösszesen csak 1 cégbe fektessünk be, kicsit meg kell változtatni a Kelly-stratégia használatát.

Általánosan azt lehet mondani, hogy a kockázatos eszközök loghozamainak lapultsága a normális eloszlásnál magasabb (leptokurtózis), és ennek a stilizált ténynek eleget tesz, ha a loghozamok negyedik momentuma (tehát a kurtózis, lapultság) nem létezik, ami a 4 vagy annál kisebb szabadságfokú Student-féle t-eloszlásnál fennáll például. Breymann *et al.* (2003) Student-féle t-eloszlást illesztettek a loghozamokra, és becsült szabadságfokuk 4 körül volt, míg Platen & Heath (2009) 3.64-et becsültek ugyanerre.

A kutatás keretei közt feltételezem, hogy a befektető 30 éven keresztül minden munkanap egy cégre fogad a Kelly-kritérium szerint továbbá a rákövetkező napon a likviden maradt vagyonát használja tovább. Feltételezem továbbá, hogy a realizált hozam sikeres felvásárlás esetén 20.98% várható értékkel rendelkező Student-féle 4 szabadságfokú t-eloszlás szerint alakul (jelöljük ezt $p_{realizált}$-ként). Sikertelen befektetés esetén a realizált hozam -1.62% várható értékkel rendelkező Student-féle 4 szabadságfokú t-eloszlást követ ($q_{realizált}$). Mivel a felvásárlás megvalósulása idejének eloszlásáról nincs információm, így azt 3 éven keresztül egyenletesnek tételeztem fel. Minden fogadást ezért 1.5 év elteltével számolunk el, ahol sikeresség esetén az akkor kockáztatott érték $1 + p_{realizált}$-szorosa íródik a számlához (és a kettős könyvelés elve alapján 1-szeres részvény íródik le), ellenkező esetben pedig a kockáztatott vagyon $1 - q_{realizált}$-szorosa. A tét megrakása és a hozam elszámolása közti 1.5 évben a likviden maradt vagyonunkat használjuk tovább, és minden nap fogadunk cégekre. Ez azzal is jár, hogy az első 1.5 évben a stratégia nem generál se bevételt, se költséget, már ami az összvagyont illeti.

Az optimális, a vagyon növekedési rátáját maximalizáló, hosszútávú stratégia tehát a következő:

1. Vizsgáljunk meg egy tetszőleges amerikai céget, amelynek az adatai megtalálhatóak a Bloombergen. Töltsük le a következő adatokat a céghez:
   (a) Összes adósság/Saját tőke
   (b) Összértékesítés/Összes eszköz
   (c) Eszközarányos árbevétel
   (d) EBIT/Árbevétel
   (e) Készpénz/Összes eszköz
   (f) Cash Flow/Részvények száma
   (g) Könyv szerinti érték/Piaci kapitalizáció
2. A céget csatoljuk hozzá a tanulómintához (még sztenderdizálás előtt), majd ott végezzük el a winsorizálást. Amennyiben az új vállalat extrém értékekkel rendelkezik, úgy ki lesz cserélve az összes extrém változója az oda tartozó (95%-os vagy 5%-os percentilishez tartozó) vektorral, de amennyiben nem, úgy benne marad az új, összecsatolt mintában. Végezzük el újra a többszörös imputálást az egyesített mintán.
3. A legjobb - keresztvalidációval optimalizált hiperparaméterekkel rendelkező - SVM modellel becsültessük meg a felvásárlás valószínűségét.
4. Amennyiben ez a valószínűség nagyobb, mint 7.17%, úgy fektessük a cégbe a vagyonunk Kelly által javasolt részét :$f = \frac{p \cdot (\frac{20.98}{1.62} - 1) - 1}{\frac{20.98}{1.62}}$, ahol p a felvásárlás valószínűségét jelöli (0.0.5 egyenlet).
5. A sikeres felvásárlást követő hónapban adjuk el a részvényeinkent, és a befektetést követő 3. évben zárjuk le a pozicionkat mindenképpen.

Végezzük el ezt a stratégiát a validáló halmazon! Induljunk 1 millió forintból, és 30 éven keresztül minden nap döntsünk egy, a validáló halmazból random kiválasztott cég sorsáról. Az online függelékben található kelly.png ábrát tekintve látszik, hogy a likvid vagyon maximum értékei és az összvagyon exponenciálisan növekedik az időben, ami a Kelly-kritérium használata miatt van, ugyanis az a vagyon növekedési rátáját maximalizálja. A likvid vagyonban való le- és fel ugrások annak köszönhetőek, hogy viszonylag gyakran volt olyan cég, amire az SVM elég nagy felvásárlási valószínűséget tulajdonított, így a vagyonunk majdnem egészét olyankor befektettük az üzletbe. Az is látszik az alsó két ábrára tekintve, hogy a szimplán random benchmark stratégiához[11] képest a dolgozatban levezetett modell sokkal jobban teljesített. A napi növekedési ráta nagyjából stabilizálódott mindkét esetben, és ez az SVM modell esetében körülbelül 0.41 bázisponttal magasabb volt (minden nap!) modell. A stratégia 30 év alatt 1 millió forintból 25 millió forintot generált, ami nagyjából **11.326%**-os hosszú távú évi többlethozamnak felel meg.

Összehasonlításképpen a Dow Jones tőzsdeindex, ami Amerika legfontosabb 30 vállalatát foglalja magában 1988. március 25-én 1 978.95 dollár volt, míg 2018. március 26-én ugyanez 24 202.60 dollár volt, ami évi 8.705%-os hozamnak felel meg a Dow Jones indexet tekintve. Ha feltételezzük, hogy a felvásárolt és fel nem vásárolt cégek bétája átlagosan ugyanakkora, akkor a stratégia bétája egy, így a Kelly-stratégia alfája $\alpha = 11.326\% - 8.705\% = 2.621\%$

A stratégia eredménye nagyban függ a Palepu által közölt számoktól, ami a modell egy fontos limitációját adja. A stratégia használhatóságának megállapítása végett ezért elvégeztem egy szenzitivitás tesztet is. A teszt során feltettem, hogy az imaginárius befektető a Kelly-arányok kiszámolásánál csak az előzetes várható hozamokat ismerte és használta, tehát $\frac{c_1}{c_2} = \frac{20.98}{1.62}$. Azonban a hozamok realizációjánál azok nem ilyen várható értéket követtek, hanem 20%-től 10%-ig a pozitív esetben, és -10%-től -5%-ig a negatív esetben. A realizált hozamok ebben az esetben is Student-féle 4 szabadságfokú t-eloszlást követtek az említett, előzetesen ismeretlen várható értékekkel. A szenzitivitás teszt eredménye alapján a lehető legrosszabb szcenárió

[11] amikor minden céghez egyenletesen sorsolunk felvásárlási valószínűséget, és arra alkalmazzuk a Kelly-formulát.

bekövetkezte esetén is, a 30 év alatt megtermelt összvagyon 1 millió forintból 2 977 125 forint, tehát sose fordul a stratégia veszteségesbe[12].

8. Összefoglalás

Kutatásom célja az volt, hogy a Support Vector Machine módszerének alkalmazásával minél pontosabban megbecsüljem a vállalatok felvásárlási valószínűségét, és a modellre támaszkova adjak egy kereskedési stratégiát. A pénzügyi szakirodalom alapján választottam ki a modellbe bevonandó magyarázó változókat. Az elemzés során használt minta elemszáma lényegesen nagyobb volt, mint a korábbi munkákban. A 26 001 rekordból álló adatbázis jellentős mennyiségű hiányzó értéket tartalmazott. Ezt a többszörös imputálás módszerével kezeltem. A kilógó értékeket winsorizálással és vertikális outlier-szűréssel filtereztem ki.

A kutatásban a minta megtisztítása után az SVM Gauss-kernel változatának hiperparamétereit ötszörös keresztvalidációval, és huszonötszörös bootstrap móddszerrel is optimalizáltam rácsmenti keresés mellett. A tanító halmazon elért nagyobb AUC-mérték a keresztvalidált SVM modell használatát implikálta. Az Out-Sample AUC mérték 93.1%, ami rendkívűl nagynak mondható. A kutatásban részletes stratégiát adtam hosszú távú kereskedést feltétezve. A Kelly-kritériumot applikáltam a vagyon növekedési rátájának maximalizálása végett, amivel évi több, mint 11%-os hozamot generált a stratégia, 30 éves befektetési időtávot tekintve. A stratégia sikerességét a szenzitivitás-teszt is alátámasztja. Tekintve, hogy a modell szerint valóban lehetőség van a vállalati felvásárlásokra való fogadással statisztikai arbitrázsra, úgy ez az eredmény a vállalati felvásárlási piac hatékonyságának közepesen erős változatát tagadja.

Irodalomjegyzék

BARNES, PAUL. 1996. The regulation of insider dealing in the UK: some empirical evidence concerning share prices, merger bids and bidders' advising merchant banks. *Applied Financial Economics*, **6**(4), 383–391.

BREYMANN, WOLFGANG, DIAS, ALEXANDRA, & EMBRECHTS, PAUL. 2003. Dependence structures for multivariate high-frequency data in finance.

CRAWFORD, DEAN, & LECHNER, THOMAS A. 1996. Takeover premiums and anticipated merger gains in the US market for corporate control. *Journal of Business Finance & Accounting*, **23**(5-6), 807–829.

DANIS, ILDIKÓ. 2012. Az adathelyettesítés modern technikája- „Multiple Imputation (MI)". *Alkalmazott Pszichológia*, 65–70.

EFRON, BRADLEY, & TIBSHIRANI, ROBERT J. 1994. *An introduction to the bootstrap*. CRC press.

FAMA, EUGENE F. 1970. Efficient market hypothesis: A review of theory and empirical work. *Journal of Finance*, **25**(2), 28–30.

FOGELBERG, GRAEME, LAURENT, CLINTON R, & MCCORKINDALE, DEREK G. 1978. The usefulness of published financial data for assessing the takeover vulnerability of companies in New Zealand. *New Zealand Economic Papers*, **12**(1), 184–206.

HARRIS, ROBERT S, STEWART, JOHN F, GUILKEY, DAVID K, & CARLETON, WILLARD T. 1982. Characteristics of acquired firms: fixed and random coefficients probit analyses. *Southern Economic Journal*, 164–184.

HSU, CHIH-WEI, CHANG, CHIH-CHUNG, LIN, CHIH-JEN, *et al.* 2003. A practical guide to support vector classification.

JENSEN, MICHAEL C, & RUBACK, RICHARD S. 1983. The market for corporate control: The scientific evidence. *Journal of Financial economics*, **11**(1-4), 5–50.

KELLY JR, JOHN L. 2011. A new interpretation of information rate. *Pages 25–34 of: The Kelly Capital Growth Investment Criterion: Theory and Practice*. World Scientific.

LEWELLEN, WILBUR G. 1971. A pure financial rationale for the conglomerate merger. *The Journal of Finance*, **26**(2), 521–537.

MIN, JAE H, & LEE, YOUNG-CHAN. 2005. Bankruptcy prediction using support vector machine with optimal choice of kernel function parameters. *Expert systems with applications*, **28**(4), 603–614.

PALEPU, KRISHNA G. 1986. Predicting takeover targets: A methodological and empirical analysis. *Journal of accounting and economics*, **8**(1), 3–35.

PEAT, MAURICE, & STEVENSON, MAXWELL. 2008. Predicting Australian takeover targets: a logit analysis. *In: The 6th INFINITI Conference on International Finance*.

PETERS, OLE, & GELL-MANN, MURRAY. 2016. Evaluating gambles using dynamics. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, **26**(2), 023103.

PLATEN, ECKHARD, & HEATH, DAVID. 2009. *A Benchmark Approach to Quantitative Finance (Springer Finance)*. Springer. p85.

REGE, UDAYAN P. 1984. Accounting Ratios to Locate Take-Over Targets. *Journal of Business Finance & Accounting*, **11**(3), 301–311.

RUBIN, DONALD B. 1987. Comment. *Journal of the American Statistical Association*, **82**(398), 543–546.

STEVENS, DONALD L. 1973. Financial characteristics of merged firms: A multivariate analysis. *Journal of Financial and Quantitative Analysis*, **8**(2), 149–158.

THORP, EDWARD O. 2011. Optimal gambling systems for favorable games. *Pages 61–80 of: THE KELLY CAPITAL GROWTH INVESTMENT CRITERION: THEORY and PRACTICE*. World Scientific.

TIKHOMIROV, VLADIMIR M. 1996. The evolution of methods of convex optimization. *The American Mathematical Monthly*, **103**(1), 65–71.

VAPNIK, VLADIMIR. 2013. *The nature of statistical learning theory*. Springer science & business media.

[12] lásd: függelék, pozneg.png, 3dplot.png