

Pénzügyi big data szerkesztése a piacvezető Q nyelv segítségével

KOVÁCS ANDRÁS

KONZULENSEK: TIMOTITY DUSÁN, DUDÁS ÁKOS

Cél

- ▶ 2.5 TB-os adatbázis feldolgozhatóvá tétele (Tőzsdei adatok)
- ▶ Kereskedési adatok elemzéséből viselkedésminták azonosítása
 - ▶ Véletlenszerű
 - ▶ Belfentes
 - ▶ Heurisztikus
- ▶ (Vizualizáció)

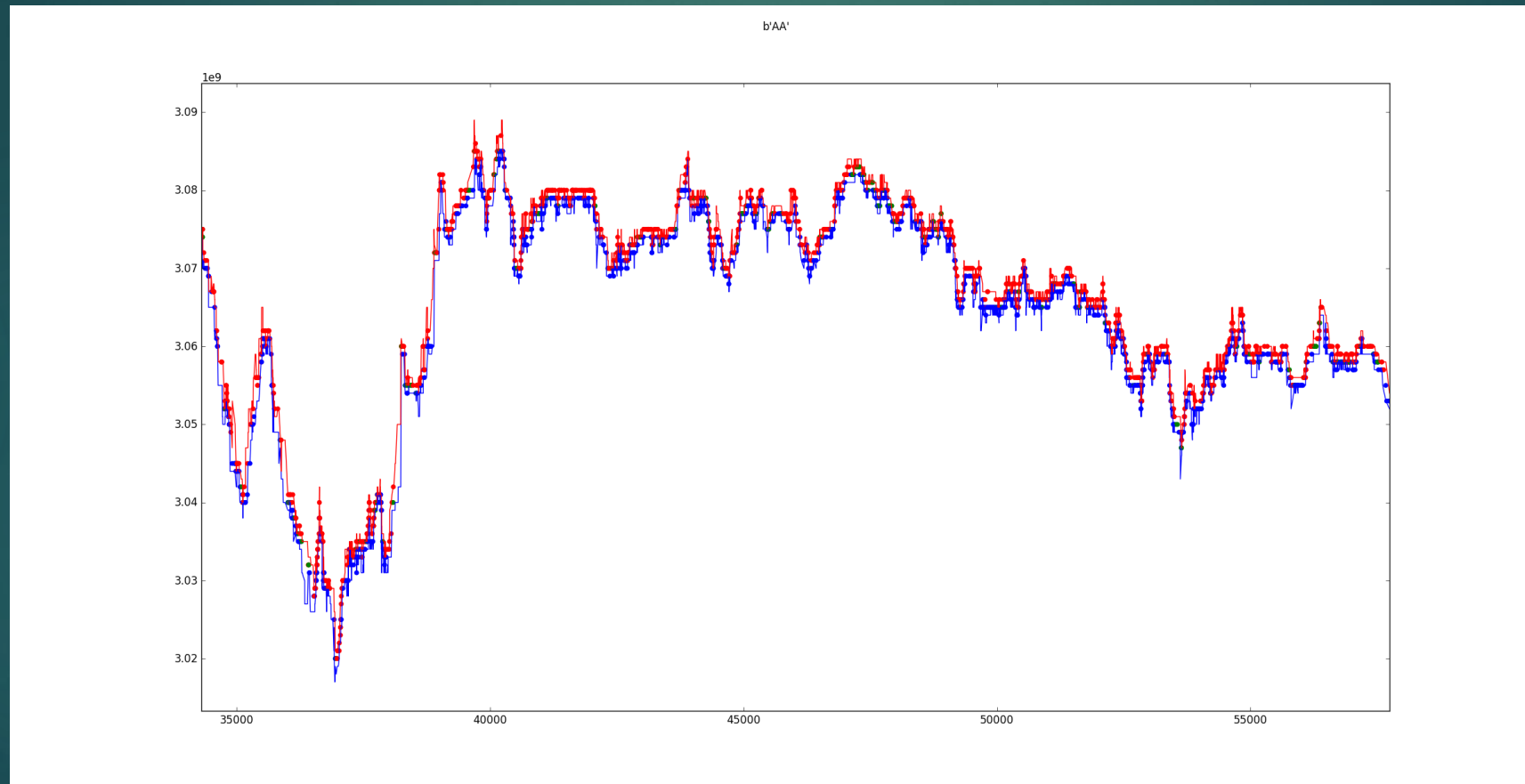
Az adatbázis I.

DATE File	CT Index	CT Binary File (Tyyyyymm.x.BIN)	CQ Binary File
<div><div>PK</div><div>TDATE</div></div> <div>CQIDXB</div> <div>CQIDX</div> <div>CTIDXB</div> <div>CTIDX</div> <div>DISK</div> <div>DATAFMT</div>	<div><div>PK</div><div>SYMBOL</div></div> <div>TDATE: yyyyymmdd</div> <div>BEGREC: start row index</div> <div>ENDREC: end row index</div>	<div>TTIM: Seconds since midnight</div> <div>PRICE: 8 implied digits</div> <div>SIZ: number of shares</div> <div>TSEQ: (MDS) sequence number if NYSE</div> <div>G127: rule 127 and stoped trade indicator</div> <div>CORR: correction indicator</div> <div>COND: condition indicator</div> <div>EX: which exchange</div>	<div>QTIM: Seconds since midnight</div> <div>BID: 8 i.d.</div> <div>OFR: 8 i.d</div> <div>QSEQ: (MDS) sequence number if NYSE</div> <div>BIDSIZ: *100 share units</div> <div>OFRSIZ: *100 share units</div> <div>MODE: quote condition</div> <div>EX: which exchange</div> <div>MMID: Identifies the NASDAQ market maker for each NASD quote.</div>
	CQ Index		
	<div><div>PK</div><div>SYMBOL</div></div> <div>QDATE: yyyyymmdd</div> <div>BEGREC: start row index</div> <div>ENDREC: end row index</div>		

Az adatbázis II.



Tradek osztályozása: Lee-Ready algoritmus I.



Architektúra

- ▶ KDB+ server
 - ▶ Nagy teljesítményű adatbázis szerver
 - ▶ Q-ban és SQL-szerű nyelven programozható
- ▶ Python script
 - ▶ Futhat ugyanazon a gépen, mint KDB+ server
 - ▶ Fő feladata a Lee-Ready végrehajtása
 - ▶ Vizualizáció (matplotlib)
 - ▶ Exportálás (pl: csv)
- ▶ Wolfram Mathematica
 - ▶ Szimbólikus algebra
 - ▶ Maximum likelihood illesztés
 - ▶ Extra adatok egyszerű beszerzése (S&P500 index)

A modell I.

- ▶ Véletlenszerű műveletek
 - ▶ Likviditási probléma
 - ▶ Hosszú távú befektetés
- ▶ Heurisztikus
 - ▶ Piaci hozam alapján dönt
 - ▶ Veszteség után, inkább kockáztat
- ▶ Bennfentes
 - ▶ Belső pozitív vagy negatív információ alapján vesz/ad el

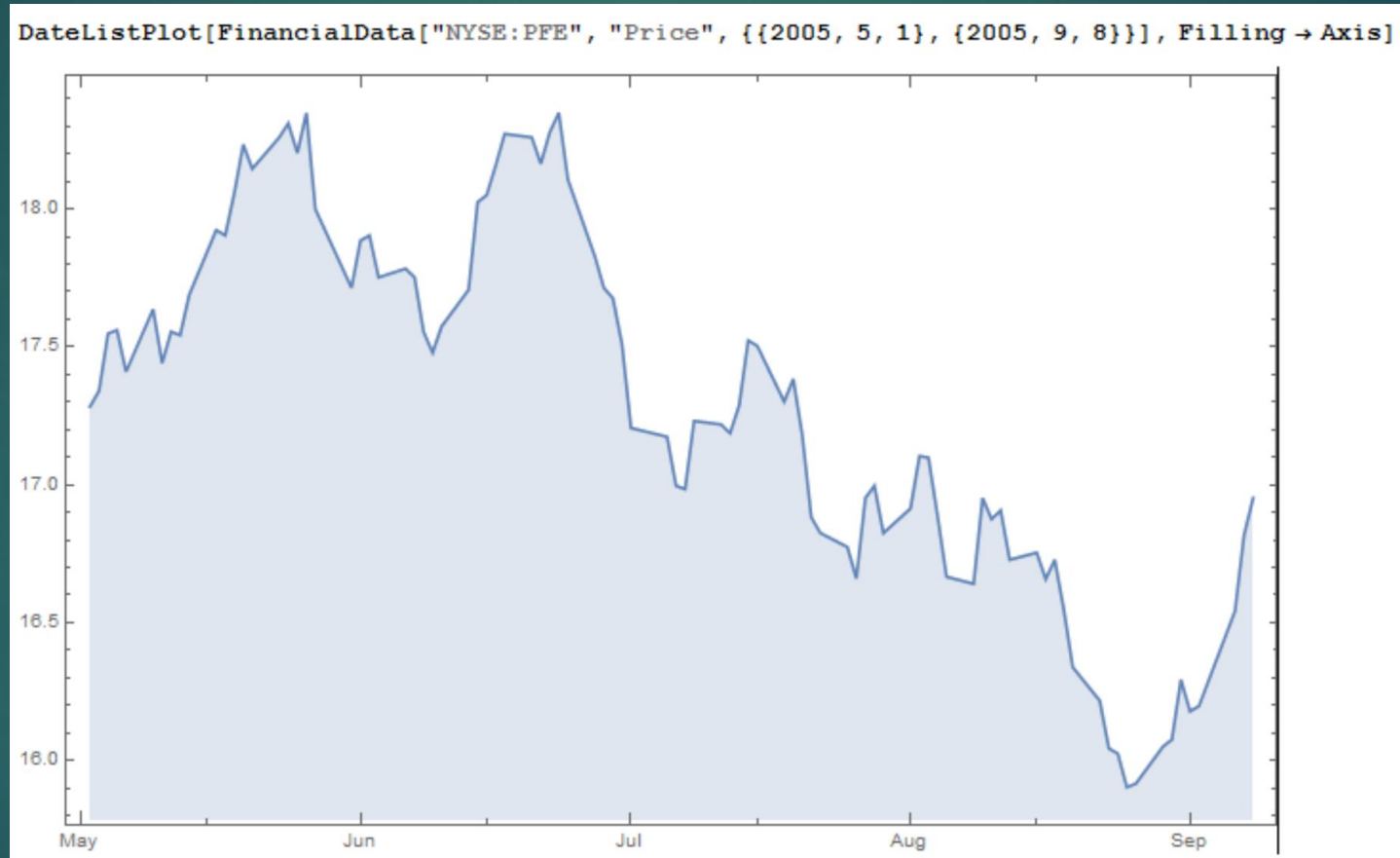
A modell II.

- ▶ Hírmmodell
 - ▶ Alpha: Hír bekövetkezésének gyakorisága
 - ▶ Delta: A hír iránya (0:pozitív 1:negatív)
- ▶ Csoportmodell
 - ▶ Minden tag egymástól függetlenül dönt
 - ▶ Valamilyen várható értékkel
 - ▶ Poisson eloszlás

Adat I.

SYMB	DAYS	DIR	RB	HB	RS	HS	INF
IBM	0.54	0.51	2388.76	139.73	2120.19	293.45	542.99
JNJ	0.44	0.48	2264.85	52.95	1725.37	45.16	525.07
PFE	0.61	0.88	2659.63	91.80	2504.07	11.47	467.54

Adat II.



Köszönöm a figyelmet