

From Buzz-words to real work

הקדמה

בשנים האחרונות יש עניין רב ביישום אנליטיקה מתקדמת בתהליכים עסקיים בארגון. אלגוריתמים חדשים מתפרסמים חדשות לבקרים ואפילו נוצר תפקיד שזהו כל מטרתו – מדען הנתונים הל-הוא ה Data scientist – בעולם ייצור החשמל נעשה כבר שנים רבות שימוש בחיזוי סטטיסטי של הצריכה העתידית. תחזית זו היא חלק מתהליך תכנון הייצור אשר מתבסס על התחזית הסטטיסטית בשילוב מידע נוסף כגון יכולת הייצור הזמינה כדי לבנות את תוכנית הייצור. תהליך זה הוא בליבת הפעילות של כל ארגון העוסק בייצור או סחר בחשמל והוא חייב להשתלב בצורה הטובה ביותר בתפקודה היום-יומי של החברה.

באופן טבעי, כל דיון על חיזוי מתנקז לנושא דיוק התחזיות. ההתפתחות המואצת של עולם מחקר הנתונים כיום מייצרת עשרות ומאות שיטות חיזוי שונות שכל אחת מהן מבטיחה להשיג את הדיוק הטוב ביותר ולהפוך את כל השיטות האחרות למיותרות עם גורל דומה לזה של הפתיליה, הסיפולוקס והווקמן.

אין ספק שדיוק התחזיות הוא דרישה חשובה אבל מטרת הרצאה זו היא להציג מספר היבטים נוספים שיש לתת עליהם את הדעת בעת בניית תהליך חיזוי. המעבר מדיבורים ורעיונות לתהליך מובנה ויציב שיכול לתפוס את מקומו הראוי בין כל יתר התהליכים התפעוליים הקריטיים לפעילות הארגון איננו טריוויאלי וישנם מגוון תחומים ובעיות פוטנציאליות שיש לתת עליהן את הדעת.

הגדרת חישוב הדיוק

הסוגיה הראשונה היא הגדרה נכונה של חישוב הדיוק התחזית. קיימים עשרות מדדים שונים של דיוק בספרות המקצועית ולכל אחד מהם יתרונות וחסרונות. אנחנו צריכים לבחור את המדד המדמה בצורה הטובה ביותר את הדרישות העסקיות של תהליך החיזוי. לדוגמא, אם הארגון נדרש לתחזית חצי-שעיתית אין משמעות למדידת טיב החיזוי ברמת סה"כ יומי ולהיפך. חלק ממדדי החיזוי מכילים הטיה מובנית שיש לקוח בחשבון שבעת בחירת חישוב כלשהו המייצג את טיב החיזוי על פני חישוב אחר.

סוגיה נוספת היא הרמה בה מחושב הדיוק. האם הארגון מחויב לרמת דיוק מסוימת ברמת כל מונה בודד, ברמת לקוח (הכולל מספר מונים) או ברמת סה"כ הצריכה בלבד? ככל שעולים ברמת הקיבוץ (ממונה ללקוח ולסה"כ) מבחינה סטטיסטית קל יותר להשיג רמת דיוק טובה יותר.

גם תקופת הבדיקה עליה נבחן טיב החיזוי משחקת תפקיד. אם אנחנו רוצים לבחון את טיב האלגוריתם על פני תקופה יציבה יחסית או דווקא לראות איך הוא מתנהג בתקופה של שינויים חדים למשל בעונת המעבר או בתקופות של חגים ואירועים מיוחדים.

נתוני מקור

בנוסף לנתוני הצריכות בפועל תהליך החיזוי דורש גם נתונים נוספים שיש לעבד ולהכין עבורו. חלק מהנתונים, למשל נתוני מזג אויר או היום בשבוע, הם משתנים מסבירים נדרשים לאלגוריתם החיזוי עצמו אבל ישנם נתונים נוספים הנדרשים לתהליך החיזוי מההיבט התפעולי. לדוגמא, תהליך החיזוי צריך לדעת איזה מונים הם פעילים בטווח התחזית כדי שלא נחזה מונים של לקוחות שכבר עזבו אותנו או שנשכח לחזות מונים אשר צפויים להצטרף לארגון בתוך אופק התחזית. דוגמא נוספת לנתונים מסוג זה היא תקופות שעון הקיץ הדורשות עדכון במידה והיה בהן שינוי.

בעיה של נתוני המקור היא הגעה של נתונים חריגים או אי הגעה של נתונים כלל עקב תקלה או קשיים אחרים. תהליך החיזוי חייב לכלול מנגנונים להתמודדות עם מקרים שכאלו כדי למנוע "השתוללות" של התחזיות המתקבלות. לדוגמא, תחזיות מזג האוויר מגיעות, לרוב מספק חיזוני אולם במידה ובריצה מסוימת אין נתוני תחזית מזג אויר זמינים מכל סיבה שהיא תהליך חיזוי שתוכנן היטב יידע לעשות שימוש בנתונים תחליפיים כגון טמפרטורה ממוצעת, תחזית מזג אויר ישנה יותר או כל נתון תחליפי אחר שיכול לאפשר לתהליך החיזוי לסיים בהצלחה גם אם לא ברמת הדיוק המקסימלית האפשרית.

יצבות וגמישות האלגוריתם

אלגוריתם החיזוי הוא לב התהליך וכל יתר החלקים של תהליך החיזוי נבנים כדי לתמוך בו. כאשר אנחנו בוחרים אלגוריתם חיזוי יש לקחת בחשבון מספר שיקולים, שחלקם אף סותרים אחד את השני. האלגוריתם צריך להיות מספיק גמיש כדי להתמודד עם מורכבות חיזוי צריכת החשמל אבל מצד שני מספיק יציב כדי לתת תוצאות אמינות ריצה אחרי ריצה ללא קפיצות ותנודות חדות מידי כתוצאה משינויים מינוריים במשתנים המסבירים.

האלגוריתם צריך להיות מספיק מורכב כדי להכיל את קשרי הגומלין המורכבים של צריכת החשמל מחד ומאידך הוא צריך לרוץ מספיק מהר כדי שריצת החיזוי תסתיים בחלון הזמן שיש בין קבלת נתוני הקלט (צריכות האתמול, נתוני טמפרטורה וכו') ולבין המועד המאוחר ביותר בו נדרשת התחזית המעודכנת וברמת הפירוט הנדרשת לארגון (מונה/לקוח/סה"כ, חצי-שעתי/שעתי/סך יומי וכו').

מודולריות

בנייה מודולרית של תהליך החיזוי ע"י שורת שלבים עם מינימום תלות בין שלב אחד לשני חיונית לתחזוקה שוטפת של התהליך לאורך זמן. האפשרות להחליף את אלגוריתם החיזוי באלגוריתם טוב יותר ללא צורך בבנייה מחודשת של התהליך כולו נדרשת לשמירה על תהליך חיזוי איכותי לאורך זמן.

מעבר לכך, בנייה מודולרית של התהליך תקל עלינו במגוון רחב של תסריטים כגון החלפת ספק נתוני תחזית מזג האוויר, עדכון או החלפה של מערכת ניהול הלקוחות ועוד.

ניהול תהליך

כמובן שלכל תהליך קריטי נדרש מנגנון בקרה ומעקב. מנגנון הבקרה צריך להשתלב בתשתית וביכולות של הארגון כדי לוודא שיעבוד ביעילות ולכן אין פתרון אחד המתאים לכולם.

מנגנון הבקרה צריך לאפשר מעקב הן אחרי התפעול של תהליך החיזוי – התחלת הריצה, הגעת כל הקלטים הנדרשים וסיום תקין של הריצה – והן בקרה על איכות התחזית המתקבלת – בדיקת סבירות ע"י השוואה לתחזית קודמת או השוואה בין מספר אלגוריתמים שונים.

המרכזיות של תהליך החיזוי עבור הארגון דורש גם הבנייה שלו כך שהתהליכים הנוספים בארגון התלויים בחיזוי יוכלו להמשיך לעבוד גם במידה ותהליך החיזוי עצמו לא סיים בהצלחה במשך זמן מסוים. אם הארגון נדרש לתחזית שוטפת של יום אחד קדימה אז רצוי שכל ריצת חיזוי תפיק תחזית של שבוע קדימה וכך גם במקרה חירום עדיין יהיו בידי הארגון מספר תחזיות שאומנם הן לא הכי מדויקת שיש אבל בהחלט מספקות בסיס כלשהו לעבוד איתו עד לתיקון התקלה בתהליך החיזוי

היקף הבקורות צריך גם להתאים לידע של בעל התפקיד בארגון אשר אמור לעקוב ולנהל את תהליך החיזוי. במידה ובארגון יש מדען נתונים עם הידע הסטטיסטי נדרש להתערבות בתהליך אז הבקורות יהיו ספציפיות ומעמיקות יותר. בארגונים אחרים הבקורות יספקו מידע כללי יותר על הצלחה או כישלון של ריצת החיזוי ובמידת הצורך הארגון ידע לפנות למומחה מקצועי חיצוני לביצוע ניתוחי העומק הנדרשים.

סיכום

כאשר ארגון נמצא בשלבי הקמה של תהליך חיזוי עליו לקחת בחשבון מגוון רחב של דרישות נוספות מעבר לדיוק של אלגוריתם החיזוי שעל הפרק. אומנם האלגוריתם הוא לב התהליך אך ישנם שלבים נוספים אשר חיוניים לבנייתו של תהליך חיזוי מוצלח שיתרום את המרב להצלחתו העסקית של הארגון כולו